



**HAL**  
open science

## A Novel, Network-Based Approach to Assessing Romantic-Relationship Quality

Blandine Ribotta, Miguel Alejandro Silan, Olivier Dujols, Ronan Bellemin,  
Hans Ijzerman

► **To cite this version:**

Blandine Ribotta, Miguel Alejandro Silan, Olivier Dujols, Ronan Bellemin, Hans Ijzerman. A Novel, Network-Based Approach to Assessing Romantic-Relationship Quality. *Perspectives on Psychological Science*, 2024, 10.1177/17456916231215248 . hal-04926159

**HAL Id: hal-04926159**

**<https://hal.science/hal-04926159v1>**

Submitted on 3 Feb 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **A Novel, Network-Based Approach to Assessing Romantic-Relationship Quality**

The CORE Lab<sup>1</sup>

*This manuscript is forthcoming in Perspectives on Psychological Science.*

**Author Note.** The OSF page with both projects can be found at <https://osf.io/ye2u4/>.

---

<sup>1</sup> Authors in alphabetical order by last name: Olivier Dujols, Université Grenoble Alpes; Hans IJzerman, Annecy Behavioral Science Lab, Université Grenoble Alpes, & Institut Universitaire de France; Bastien Paris, Université Grenoble Alpes; Blandine Ribotta, Université Grenoble Alpes; Miguel Alejandro Silan, University of the Philippines Dilman, Université Lumière Lyon 2, & Annecy Behavioral Science Lab; Caspar van Lissa, Tilburg University. To whom correspondence should be addressed: Hans IJzerman ([hans@absl.io](mailto:hans@absl.io)).

**Abstract**

How should romantic relationship quality be approached psychometrically? This is a complicated theoretical and methodological challenge, which the authors start to address through three studies. In Study 1a, the authors identified 25 distinct romantic relationship categories among 754 items from 26 romantic relationship quality instruments, with a weak Jaccard index (0.38), indicating that the scales' item content was very heterogeneous. Study 1b then demonstrated limited structure validity evidence in 43 scale development-validation articles of 23 of these 26 instruments. Finally, Study 2 surveyed 587 French-speaking participants in a romantic relationship on romantic relationship quality. Applying a network-based model, four dimensions were identified, with three items being central to relationship quality. The inferences were mostly limited to French-speaking, monogamous, heterosexual women. To resolve challenges detected in the literature, the authors recommend a multi-country qualitative approach, more diverse sampling, better definitions of romantic relationship quality, and a dynamical systems approach to measuring romantic relationship quality. We provide a manual for the new scale on our OSF page: <https://osf.io/qj69p>.

**A novel, network-based approach to assessing romantic-relationship quality**

Romantic relationships are a fundamental part of our lives and play a significant role in our health and well-being (Fellows et al., 2016; Proulx et al., 2007; Robles et al., 2014; Wahyuningsih et al., 2020). However, assessing the quality of a romantic relationship is highly complicated and, there is no consensus in the definition of the concept. The absence of a well-defined construct potentially signals that we may not be measuring the same relationship-quality construct across studies. As a result, the interchangeable use of instruments that measure very different aspects of relationship quality could pose a threat to overall understanding of romantic relationships.

Furthermore, even when there is a clear understanding and a consensus of *what* to measure, it is important to understand *how* to measure. Central to being able to understand any concept in the psychological literature is construct validation of the measures assessing these concepts. Despite this general understanding, recent articles have indicated the poor state of construct validation in the field (Flake & Fried, 2020; Fried & Flake, 2018; Hussey & Hughes, 2020) and a common practice of the underreporting of validity evidence (Flake et al., 2017).

For these reasons, in Studies 1a and b, we conducted a comprehensive search of the instruments assessing romantic-relationship quality and created a comprehensive database of scales assessing romantic-relationship quality. We then assessed the quality of romantic relationship measurements in two ways: we first explored item-content overlap across these scales that shows what the scales are measuring. Second, we examined the extent to which researchers reported the structure validity of these instruments (“how” the relationship quality is measured). As part of our assessment, we also identified the scale’s country of origin. In Study 2, in a sample of 587 French-speaking participants, we provided an alternative perspective for the

conceptualization and assessment of romantic-relationship quality using network analysis. There we identified the complex associations between different indicators of romantic-relationship quality and identified the central items of romantic-relationship quality.

### **Romantic-relationship quality and threats to its assessment**

What is relationship quality? In the research literature, authors conceptualize it in many ways, but no clear consensus has been established. Romantic-relationship quality is often described as the subjective evaluation of the relationship by either partner at one point in time (see, for instance, Glenn, 1990). But romantic-relationship quality is often used interchangeably with the term *relationship adjustment*, referring to a dynamic process that consists of events, circumstances, and interactions moving the couple through a continuum of adjustment (Spanier & Cole, 1976). Two other dimensions that are typically deemed of importance and sometimes used interchangeably with relationship quality are satisfaction and happiness in the relationship (Fincham & Bradbury, 1987; Norton, 1983), while intimacy, conflicts, cohesion, consensus, communication, and thoughts of separation are regularly employed as subcomponents of relationship quality (Delatorre & Wagner, 2020).

More recent approaches have focused on goal-pursuit of partners (Fowers & Owenz, 2010), commitment and attachment (Stanley et al., 2010), and ethical responsiveness (Galovan & Schramm, 2018). However, at times these components of the quality of a romantic relationship are regarded as *predictors* instead (Karney & Bradbury, 1995). For instance, the frequency of conflicts can predict overall relationship quality (although relationship quality can also predict conflict frequency instead; Kluwer & Johnson, 2007). The conceptual and structural models differing across studies increases the risk of *interpretational confounding*, making any interpretations of these different studies' results problematic (Howell et al., 2007) and will occur

more easily when researchers use the same term but employ distinct composites (Markus & Borsboom, 2013).

How has romantic-relationship quality – and how have some of these components – been measured? Since the 2000s, researchers have increasingly recommended to assess romantic-relationship quality with multidimensional instruments given its complex nature (Delatorre & Wagner, 2020; Fincham & Rogge, 2010; Fowers & Owenz, 2010; Hassebrauck & Fehr, 2002; Reblin et al., 2020). Fletcher et al. (2000), for instance, have compared the fit of several models explaining romantic relationship quality. They showed that a multidimensional model explains a greater amount of variance of romantic relationship quality when compared to a unidimensional model. Although unidimensional instruments could solve methodological challenges, such as potential intercorrelations between predictor and criterion variables (Fincham & Bradbury, 1987; Sabatelli, 1988), these unidimensional instruments cause significant variance in the relationship to not be explained (Delatorre & Wagner, 2020; Hassebrauck & Fehr, 2002).

Nevertheless, Delatorre and Wagner (2020) still observed that self-report scales currently in use in the literature to assess romantic-relationship quality are 1) predominantly unidimensional and 2) from the United States. Further, they find that marital quality measures usually only reported exploratory evidence of construct validity and provided fuzzy conceptual definitions or sometimes no definitions at all. As but one out of many examples, the Dyadic Adjustment Scale measures dyadic adjustment, which is thought to be determined by the degree of: “(1) troublesome dyadic differences, (2) interpersonal tensions and personal anxiety, (3) dyadic satisfaction, (4) dyadic cohesion, and (5) consensus on matters of importance to dyadic functioning” (Spanier, 1976, p. 17). And yet, in the Dyadic Adjustment Scale, five domains were compiled into four dimensions: satisfaction, cohesion, consensus, and affectional expression.

Items representing troublesome dyadic differences and interpersonal tensions are combined in the consensus dimensions and in the satisfaction dimension; the affectional expression dimension does not have a counterpart in its conceptual definition. Challenges to the conceptual clarity can be explained by the fact that these scales may have been constructed ad hoc or/and were based on data-driven approaches (such as exploratory analysis) and/or on authors' clinical experiences (Delatorre & Wagner, 2020). It is essential that theoretical and conceptual backgrounds are clearly defined to ensure a clear concept-measurement model mapping. Given these various potential complications, it is still unclear where the current quality of the assessment of romantic relationship stands.

### **Heterogeneity in measurement and challenges to internal structure validity**

More generally speaking, the lack of conceptual clarity is often at the heart of many measurement problems (Flake & Fried, 2020; Fried & Flake, 2018) and often becomes revealed in the presence of too many different scales (and thus items) seeking to capture the same concept. For instance, even for a concept that was thought to be as well understood as depression, 280 different scales exist to measure it (Santor et al., 2006). When Fried (2017a) conducted an analysis of the item-overlap of the seven most often-used depression scales, he found little overlap between these scales even with one of the scales being designed to assess the original DSM-5 criteria (Inventory of Depressive Symptoms; Rush et al., 1996).

The absence of well-defined psychological constructs increases the likelihood that researchers may not be measuring the same concept across different studies even as they refer to one construct (Fried, 2017a, 2017b; Visontay et al., 2019; Weidman et al., 2017) resulting in jingle-jangle fallacies (Flake, 2021; Kelley, 1927; Thorndike, 1904). When two measures that are labelled the same, but represent two conceptually different constructs (jingle), or when they are

identical, but are differently labelled (jangle), one jingle-jangles. The great variety of instruments used to capture an unclearly defined construct may therefore not be comparable in content, threatening the entire process of the relationship quality literature as a cumulative process. Thus, low item-content overlap between scales may lead to idiosyncratic results across studies and may threaten the measurement validity of a large number of studies, rendering comparisons between studies hard, if not impossible (Fried, 2017a). An assessment of the (lack of) overlap between measures and items of those measures can thus help understand to what extent there is a jingle-jangle problem.

Although conceptual clarity is necessary, it is not sufficient for construct validity. According to the 2014 Standards for Educational and Psychological Testing (American Educational Research Association, 2014), construct validation is an ongoing process that requires scale developers and users to obtain and accumulate evidence that supports the intended interpretation of scale scores for a proposed use. The argument-based approach to validation presented by Kane (1992, 2013) considers proposed interpretations and uses as valid when they are explicitly stated and supported by appropriate evidence. Claims about the validity of the interpretations and uses of test scores depend on the degree to which all the accumulated evidence supports the assumptions made in the argument (see Bandalos, 2018). The 2014 Standards have outlined five sources of validity evidence based on (1) test content, (2) response processes, (3) internal structure, (4) relations to other variables, and (5) consequences of testing. All five sources of evidence contribute to the validation process and researchers should determine which ones are most appropriate based on their focus or application.

Despite the availability of these guidelines, Hussey and Hughes (2020) have noted that our field is in the practice of underreporting validity evidence, which “leaves the field in a sticky



situation: It is currently impossible to know whether the field is facing a mere problem of underreporting (as highlighted by Flake et al., 2017) or the potentially deeper issue of hidden invalidity”. Hussey and Hughes (2020) suggested the existence of many hidden invalidities among commonly used scales in social and personality psychology. When the scales were assessed and evaluated in terms of internal structure validity (using internal consistency, immediate and delayed test-retest reliability, factor structure, and measurement invariance for age and gender groups), only 4% of these commonly used scales in social and personality psychology demonstrated good validity (Hussey & Hughes, 2020; but see Wetzel & Roberts, 2020).

In the case of measurement instruments targeting romantic-relationship quality, we expected only limited evidence of internal structure validity, high heterogeneity, and we expected considerable development work of measures to be needed. For that reason, we only took the first step to investigate reported internal structure validity of instruments assessing romantic-relationship quality.

### **Research Overview Studies 1a and 1b**

The goal of Studies 1a and b was to assess the quality of romantic relationship measurement. We did so by first creating a database of scales assessing romantic-relationship quality. We then searched for articles that developed and validated romantic-relationship quality questionnaires. Following these first two steps, we divided the first two stages of the project into two studies for conceptual clarity. In Study 1a, we examined the overlap between items between different instruments assessing romantic-relationship quality. We then narrowed our search further to find articles reporting the internal structure validity of instruments we had searched prior to starting the second study. In Study 1b, we assessed internal structure validity of the

instruments we had found in our search. We coded all items into different categories of romantic-relationship quality and extracted validity evidence for each scale, such as factor analysis (EFA, CFA),<sup>2</sup> internal consistency (alpha, omega coefficients), test-retest reliability, and measurement invariance.

### **Methods**

We conducted three literature reviews to select scales assessing romantic-relationship quality and to obtain their internal structure validity evidence. We first identified existing instruments assessing romantic-relationship quality and we selected self-reported scales based on some exclusion criteria. In Study 1a, we then estimated the item content overlap between the selected scales. We conducted a thematic content analysis to assess to which degree items assess the same construct. In Study 1b, we coded the presence of validity evidence in the literature of all scales.

#### **Search Strategies of Literature Reviews**

We conducted three literature reviews to select scales assessing romantic-relationship quality (Search 1, see Scale Selection section) and to obtain their internal structure validity evidence (Search 2a and 2b; see the Study 2 section). For Search 1, we used search terms related to romantic-relationship quality (e.g., marital quality, marriage quality) and its measurement (e.g., scale, questionnaire). For Search 2a, we used the name of the scale as search terms, as well as terms related to internal structure validity evidence, such as “reliability”, “internal consistency”, “test-retest”, “factor analysis”, and “measurement invariance”. For Search 1 and Search 2a, we used the three following databases: Scopus, Web of Science, and ProQuest (which

---

<sup>2</sup> We decided to search for EFA and CFA reports because we wanted to count how often internal structural validity evidence was reported for these scales. In the studies we report, researchers usually report internal consistency evidence for scales they use, and usually report Cronbach’s Alpha. But according to the Tau equivalence model, unidimensionality is assumed to compute Cronbach’s Alpha (whereas unidimensionality is not assumed for MacDonald’s Omega). Knowing the factor structure (provided by EFA or CFA) is required prior to investigating internal consistency, which is the reason we provide an assessment of this information.

included research coming from APA PsycArticles, APA PsycINFO, ProQuest Dissertations, Theses Global). We selected these three databases because they cover a large portion of the psychological literature. For Search 2b, we searched on Google Scholar the name of each scale. We used Google Scholar to expand the scope of the results, as the three previous database results does not include articles related to some of the scales.

All search terms, Boolean operators, and search strings used for each database are available on the OSF page (<https://osf.io/kgsz8>). We imported all records into Zotero in order to delete duplicates (same title and authors, published in the same year, or published in the same journal). We used Rayyan QCRI (Ouzzani et al., 2016), a web and mobile app for systematic reviews, to carry out the abstract, title, and full-text screening. Using Rayyan QCRI also allowed us to double-check the presence of duplicates and to remove them manually.

### **Scale Selection Process and Results**

*Process.* In April 2021, Blandine Ribotta conducted Search 1 to identify the measurement instruments used in the literature to capture romantic-relationship quality. She identified 1,387 articles and first removed duplicates ( $N = 607$ ), after which she screened titles and abstracts for relevance (leading to a further removal of 89 articles). Then, she read the full text articles, leading to a further removal of 92 articles (see Figure 1 for details regarding the exclusion criteria of the article selection process). In the end, 599 articles were initially included as eligible for further review.

Blandine Ribotta then reviewed the Method Section of these 599 articles, identifying a total of 111 instruments. She searched for the paper that originally developed the measure and decided to include the scale or not, based on 1) whether the instrument was intended to measure the quality of romantic relationships ( $N = 28$ ), 2) whether the instrument was either in French or

English or not (excluded  $N = 7$ ), and 3) whether the scale was a self-report instrument for people in romantic relationships (excluded  $N = 6$ ). Then, Miguel Silan, Olivier Dujols, and Hans IJzerman double-checked a separate 10% of the final scale selection list. The percentages of agreement with Blandine Ribotta were 58.33%, 69.23%, and 75%.

**Results.** Considering the relatively low agreement rates, we then wrote a *provisional* definition of romantic-relationship quality and decided to let that definition guide us in selecting the articles. We defined romantic-relationship quality as follows: 1) general and abstract perceptions and evaluations of either partner of their relationship (e.g., love, passion, affection, intimacy, satisfaction, adjustment, conflict management, trust, commitment, investment, willingness) or 2) that partners' behaviors can be indicators of the romantic-relationship quality (e.g., communication, conflicts, sexual interactions, proximity maintenance, the amount of time spent together, or how the partners make decisions). 3) How the relationship is influenced by or withstands exogenous factors can also reflect relationship quality (e.g., family and friend support, how they affect the relationship [not social support or network strength more generally], and how couples deal with daily-life stressors). Finally, 4) we agreed that these prior categories are not fully independent (e.g., partners can have abstract perceptions of their daily interactions or may act on the basis of their general perceptions).<sup>3</sup> As we were interested in scales assessing romantic-relationship quality as a distinct construct, scales developed to capture only a predictor or a specific component of romantic-relationship quality were not included (excluded  $N = 54$ ).

---

<sup>3</sup> Given the lack of consensus in the definition of romantic relationship quality and the low agreement rates, we needed to create a working definition that would allow us to find articles that would provide a sufficient number of articles with measurement instruments of romantic relationship quality, while still providing a relatively fair, conservative, and unbiased judgment of the literature. Our working definition was based on 1) the initial review of the articles (that provided low interrater reliability estimates), 2) the authors' knowledge of the relationship literature, and 3) feedback from a practicing relationship therapist. We acknowledge that the definition is useful for the systematic review, but will necessarily be imperfect and only provide a first step to greater theoretical coherence.

After generating a provisional definition of romantic relationship quality for the review, Blandine Ribotta redid the final narrowing down from 111 scales to 26 scales with much higher initial agreement rates (75%, 83.33%, and 100%). When disagreement still occurred, we resolved them through discussion until mutual agreement was reached. Exclusion criteria of Search 1 are detailed in Figure 1. The inclusion and exclusion lists (and their specific reasons for exclusion) of the scales can be found at <https://osf.io/8xk7j>.

In Table 1 and Figure 2, we summarized the characteristics of each instrument. Most of them originated in the United States (57.69%). Eighteen scales were multidimensional (69.23%) and the rest unidimensional. Out of the 18 scales that were multidimensional, 9 recommended summing the items for a score of relationship quality. Publication dates vary between 1959 (MAT) and 2016 (GLRSS). The number of items per scale ranges from 3 (KMSS) to 90 (CMQI). After we had found the different instruments that could assess romantic-relationship quality, we then tried to assess the item-content overlap between the scales.

### ***Study 1a Item content analysis: approach***

After selecting the scales, we estimated the item content overlap. We conducted a thematic content analysis to assess to which degree items assess the same construct. We adopted this approach from Fried (2017a), who determined depression symptom overlap among seven depression scales via a Jaccard index (0=no overlap, 1=full overlap). We estimated to what degree items were overlapping across the scales assessing romantic-relationship quality. That is, we asked whether any item in any scale overlapped with any item of any other scale for all possible combinations. We then determined the rates of idiosyncratic categories per scale. That is, do categories of items appear in one scale only?

We created our categories through an abductive logic. Blandine Ribotta carried out the first round of thematic coding and created categories based on the items and how they were worded. For example, (a) the items “My partner shows me affection” (PBSC-7) and “[My partner] Is affectionate to me” (IBM-5) were both listed under the category labelled as “Affection/Love”; while (b) the items “If I were marrying again, I would pick my present spouse” (MSS-13) and “I regret being married to my husband” (MQS-41) were both listed under the category labelled as “Commitment/Involvement”.

To ensure that the categories were accurate representations of the items, we calculated interrater agreement between authors. Three other authors (Hans IJzerman, Miguel Silan, and Olivier Dujols) cross-checked a separate 10% of the coding list (i.e., they had to independently code the items themselves by using the categories created by Blandine Ribotta or by creating new ones when necessary). For this first cross-check, full agreement (i.e., all codes are similar) with Blandine Ribotta was 47.99%, partial agreement (i.e., at least one code similar; 2nd or 3rd codes might be missing or different) was 22.90%, and no agreement was 27.77%. Disagreements were subsequently resolved through discussion until consensus was reached and a codebook that described each category in detail was generated after this first discussion.

Blandine Ribotta then applied the coding to the items based on the definitions, while the categories were modified throughout the interrater discussions. After this second revision, a second double-check was carried out, after which the mean percentages of the agreement were satisfactory, reaching 64.91% for full agreement, 16.67% for partial agreement, and 18.42% for no agreement. We considered this satisfactory, given that three raters other than Blandine Ribotta were involved and that we did two rounds of cross-checking, with at least 60% of all 754 items being double-checked.

As we expected heterogeneity among scales, we took a conservative approach during this coding process. When deciding in which category a specific item fell or combining the symptom categories into one, we differentiated items only when they clearly assessed different aspects of romantic-relationship quality. However, this qualitative analysis is subjective in nature, as no formalized or objective way exists to conduct such coding (Visontay et al., 2019). We considered items belonging to the same category, as long as they had common themes, topics, ideas, and patterns of meaning that come up repeatedly. We placed some items under multiple categories simultaneously because they fit into more than one of our categories. For example, the item “Loving him/her makes me experience the warmth of a family much more intensely” (CMQI-68) we saw as feeling the amount of love for a partner and as being connected to one’s family. We labelled such items in two categories (i.e., the category labelled as “Affection/Love” and the category labelled as “Family Life/Parenting”). The item coding process is available on the OSF (<https://osf.io/umq24>).

## Results

### Data Analysis

Following Fried (2017a), we used the Jaccard Index to estimate the content overlap. This metric is a similarity coefficient specifically for binary data that ranges from 0 (no overlap among scales) to 1 (complete overlap). We did a binary classification by collapsing specific and compound items into one category for comparison against no overlap category. As described in detail in Fried (2017a), we calculated this coefficient with the formula  $s/(u1 + u2 + s)$ , where  $s$  is the number of items two scales share, and  $u1$  and  $u2$  the number of items that are unique to each of the two scales. We used the criteria of Evans (1996) for the Jaccard Index correlation coefficient: very weak 0.00–0.19, weak 0.20–0.39, moderate 0.40–0.59, strong 0.60–0.79, and

very strong 0.80–1.0. In addition, we also calculated the rate of idiosyncratic categories per scale (i.e., categories that appear on no other scale). These analyses were conducted in R (R Core Team, 2015). We adapted the code from Fried (2017a) to our study (<https://osf.io/eud5g>).

### **Item Overlap**

The categorization process of all 754 items from 26 scales resulted in 25 categories covering a wide range of content such as love, trust, support, happiness/satisfaction, conflicts, and time spent together (see Figure 3). Two items were categorized as miscellaneous and thus were not included in the categories (“Are you competent and successful at your job and/or housework?” from the MMQ; “How happy are you with your home?” from the MMQ). In Figure 3, as one moves clockwise, one can see a decreasing number of scales that assess the different categories. Among the 25 categories, “Affection/Love” was the most common category across the scales (featured in 22 scales) followed by “Emotional Closeness/Intimacy” and “Global Happiness/Satisfaction” (both appear in 20 scales). One category is unique to a single scale, namely “Forgiveness” (which was part of the RRF). However, no scale captured all categories simultaneously, nor did one of the 25 categories appear in all scales. Occurrence of categories and their number of items are summarized in Table 2.

The mean overlap among scales was 0.38, which means there is a weak mean overlap between the scales (Evans, 1996). We present the overlap among all individual scales and mean overlap of each scale with all other 26 scales in Figure 3. Several points in Figure 4 merit highlighting:

- (1) Half of the scales (N=13) had a moderate overlap with all other scales, mean Jaccard coefficients ranging from 0.40 to 0.51. The MSS showed the highest overlap with all other scales (0.51), followed by the MQS (0.50), CMQI, QDR, RRF (0.49), CMQS



(0.48), DAS, MCLI (0.46), GRIMS (0.45), MMQ (0.44), MIRS (0.43), RMAT (0.41) and IRQ (0.40).

- (2) Eleven scales had weak overlap with all other scales, mean Jaccard coefficients ranging from 0.21 to 0.39.
- (3) The KMSS and PANQIMS exhibited the lowest mean overlap with other instruments (0.09 and 0.13 respectively).
- (4) The highest overlap among individual scales was between the CMQI and MQS (0.87; very strong overlap). This pair of scales thus encompass very similar items.
- (5) KMSS showed a zero overlap with six scales (CMQI, GLRSS, MCLI, MIRS, PBSC, and RDAS). PANQUIMS also had zero overlap with PBSC and RDAS.
- (6) When we analyzed the length of the scale and the correlation with the mean Jaccard coefficient of each scale (the mean overlap of a scale with all others) we found  $r = 0.72$  meaning the more items, the more representative the scale was of the overall population of items we had found in our search.

Table 3 summarizes to the number of categories captured per scale, the scale length and the number of idiosyncratic categories. On average, 52% of the categories were captured by any of the 26 scales. The scale that captured the most categories was the CMQI ( $N=22$ ), whereas the KMSS only captured one category. The KMSS had the lowest overlap with all other scales. In contrast, the PANQIMS had the highest rate of specific items (100%). This can be explained by the fact that the PANQIMS assesses "Global Happiness/Satisfaction" through 4 items asking to evaluate positive and negative feeling towards the spouse and the marriage considering only positive vs. negative feelings. On average, 56% of all items across scales were specific and 44% were general.

### **Study 1a Discussion**

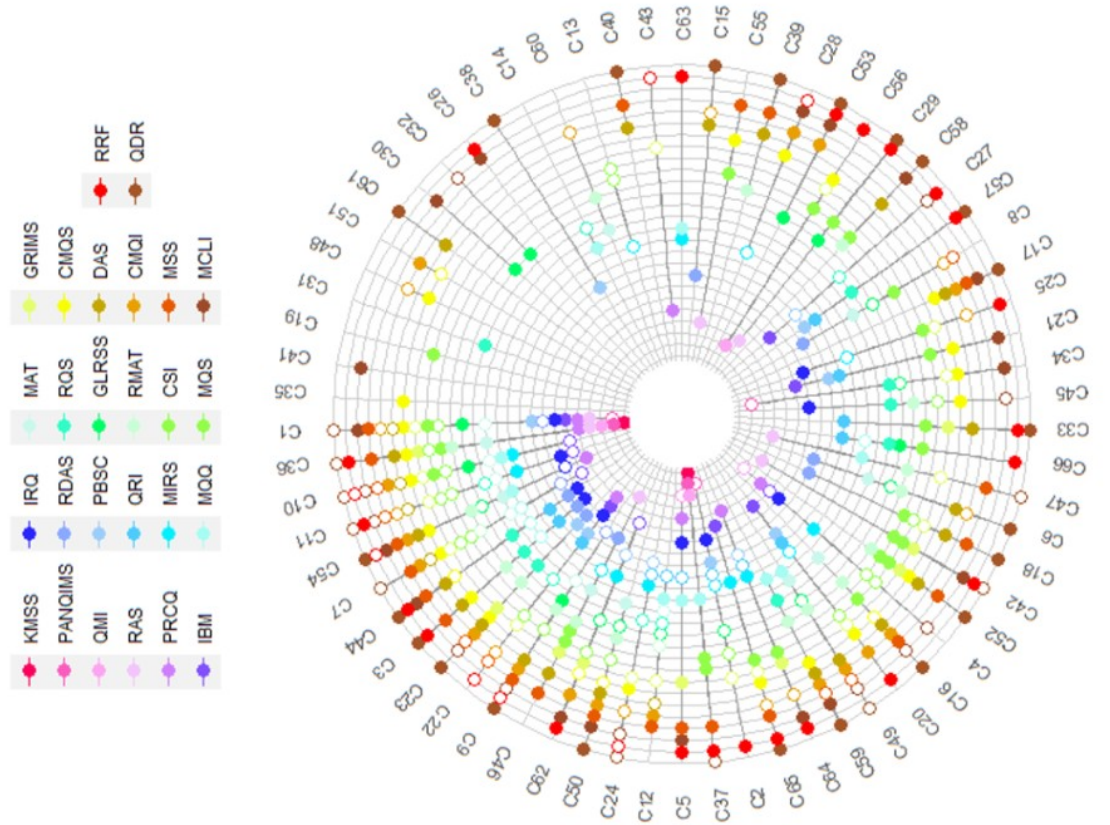
In this study, we codified 25 disparate item-content categories in 26 scales. The mean overlap of item content was weak, indicating that the categories captured by the scales were relatively heterogeneous. No scale captured all 25 item-categories and only a few scales captured a wide breadth of the item-categories, with the average item-categories captured being 13% and the average Jaccard index being 0.38. Taken together, our findings indicate a considerable heterogeneity across scales and a low degree of overlap among items.

These results were expected, given that romantic-relationship quality has no clear, consensual definition in the literature. Indeed, whether any individual scale should encompass the 25 item-categories, or only focus on one specific aspect (i.e., “Global happiness/satisfaction”) is a theoretical open question. Further, whether the item-categories should properly be thought of as predictors of relationship quality, or as components in itself is also open to debate. However, the current state of heterogeneity across scales raises the concern as to whether these scales even measure the same construct and how one can compare findings from one research project to the next. While there does not seem to be sufficient conceptual clarity, we next deliberated whether the scales as they have been constructed demonstrate sufficient internal structure validity.

#### ***Study 1b Internal structure validity: approach***

In June 2021, Blandine Ribotta conducted a second literature search to obtain the internal structure validity evidence of the scales we had found. We thus focused on validity evidence based on internal structure (i.e., relations among scale items mirror those expected from theory). We searched each scale by its name and adding search terms related to the internal structure validity such as “reliability”, “internal consistency”, “test-retest”, “measurement invariance”,

- C34 Global Compatibility
- C35 Happy/Satisfied about their home
- C36 Happy/Satisfied about their marriage/relationship
- C37 Happy/Satisfied with their partner
- C38 Hug
- C39 Kiss
- C40 Laugh together
- C41 Money owned
- C42 Need met
- C43 Passion
- C44 Quarrel (Amount)
- C45 Quarrel (Seriousness)
- C46 Regret the Marriage/Relationship
- C47 Relationship problems (Amount)
- C48 Relationships with Children
- C49 Relationships with Friends/Co-workers
- C50 Relationships with Relatives/In-laws
- C51 Role in family
- C52 Separation
- C53 Sexual Desire
- C54 Sexual Relationships
- C55 Social comparison
- C56 Stability over time
- C57 Support if needed
- C58 Time spent together (Amount)
- C59 Time spent together in Leisure time
- C60 Time spent together in Outside activities
- C61 Time spent together Working on a project
- C62 Tolerance
- C63 Touch
- C64 Trust
- C65 Understanding
- C66 Uniqueness



- C1 Affection/Love
- C2 Ambivalence
- C3 Angry/Tension
- C4 Closeness
- C5 Commitment
- C6 Compatibility in Conventionality
- C7 Compatibility in Habits/Personality
- C8 Compatibility in Interests/Activities
- C9 Compatibility in Values/Goals
- C10 Confide Feeling
- C11 Confide Personal matters
- C12 Conflict Resolution
- C13 Comfort
- C14 Competent at work
- C15 Agreement for Career decisions
- C16 Agreement for Handling finances
- C17 Agreement for Household tasks
- C18 Agreement for Major decisions
- C19 Mutual agreement for Vacations
- C20 Control/Dominance
- C21 Criticism
- C22 Dependency
- C23 Discuss together
- C24 Emotional support
- C25 Enhancement/Maintenance
- C26 Equality
- C27 Esteem support
- C28 Exclusiveness/Jealousy
- C29 Expectations met
- C30 Family/In-laws support
- C31 Financial support
- C32 Friends/Co-workers support
- C33 Future

ment

les

il

do

e

if

(A),

ilable

ed

s the

N =

cle  
or a  
arch 2

is

(27.91%). Test-retest reliability was carried out in 11 articles (25.58%, while measurement invariance evidence was reported in 3 articles (6.98%). The spreadsheet that contains validity evidence of each scale can be found at <https://osf.io/c6zww> or in Table 4.

### Study 1b Discussion

The analyses of validity reporting have revealed that the authors did not systematically report internal structure validity evidence. Coefficient alpha was by far the most common type of evidence reported, whereas the omega coefficient was not reported at all. These results underline once again the overuses of the alpha coefficient (Flake et al., 2017; Schmitt, 1996) and the slow shift from the alpha to omega coefficient in the scale development field (Dunn et al., 2013; Sijtsma, 2009). Further, the discrepancy between the scales that calculate for alpha and for those that perform factor analysis also indicate the paucity of checking for dimensional assumptions before calculating the alpha or omega coefficient (Flora, 2020; Tavakol & Dennick, 2011). Although an alpha coefficient is a “useful tool for summarizing the internal consistency of scale items as a measure of reliability” (Flake et al., 2017), the coefficient does not provide sufficient nor necessary evidence for valid measurement (Borsboom et al., 2004).

Exploratory factor analysis evidence had been reported more than confirmatory analysis to test the structure of the scales. These findings suggest that fewer scales are grounded in a theoretical basis than could be expected (Delatorre & Wagner, 2020). Three studies had tested measurement invariance, which indicates that few authors have ensured that comparisons between different groups of participants are meaningful and valid. Without measurement invariance testing, researchers cannot presume that the same construct is being measured across groups using the same measure, which is an even more pertinent issue if clinicians want to use

such scales to assess improvements of relationship quality after an intervention. Finally, test-retest reliability had been reported a few times. Taken together, our results indicate that these scales were developed and used without enough strong validity evidence.

### **Studies 1a and b Discussion**

In the first two studies, we tried to assess what relationship-quality scales measure, to what degree they overlap, and evaluate the internal structure validity evidence of these scales. In general, across the 26 relationship quality scales there is a weak overlap among item content indicating heterogeneity in what is being measured. Internal structure validity evidence mostly focuses on internal consistency measures and factor analytic models. Particularly noticeable is that 15/26 of the scales were developed in the United States and 6 out of the remaining 11 in countries where the native language is English.

### **Limitations**

One additional explanation of our overlap results is that our amount of scales and items analyzed could have increased the heterogeneity and decreased the overlap (see Fried, 2017a). In addition, the choice of the scales selected could have biased our results: The language restriction (English and French) and the fact that some articles were not always available in full-text could have limited the sample of scales. Finally, the interpretation of the Jaccard Index was under Evans's (1996) guidelines, who himself acknowledged that this was done in the absence of any well-cited guidelines. This study has some limitations regarding the inclusion criteria for the second literature search (Search 2). We limited the inclusion to two articles per scale, which reduced the number of articles analyzed. Similarly, the limited access to full-text articles, the language restriction, and the low number of articles per scale resulting from the Search 2, could have affected the sample of articles.

A priori, we had adopted a definition based on the existing romantic relationship quality literature in the interest of transparency and reproducibility. Our definition set our search terms, on which one may reasonably disagree. We are unsure how this definition impacted the selection of papers or scales, but the adoption of our definition allows other researchers to redo the selection when necessary.

### **Constraints on Generality (Simons et al., 2017)**

Note that the scales in the literature we have detected have primarily been developed amongst white, heterosexual individuals from the United States, and were all self-report, thereby severely limiting the potential generalizability of the concept of romantic relationship quality. We are unsure how the coded categories would generalize to other populations and recommend qualitative work to be done amongst other populations to better understand concerns across populations.

### **Is romantic-relationship quality really compatible with latent variable models?**

The measurement of romantic-relationship quality is perhaps more compatible with network models, representing the structure of item connections and the complexity of the categories that constitute romantic-relationship quality. In this approach, psychological constructs are emergent properties that arise out of the direct causal relations among items and are thus an appropriate alternative to rethink romantic-relationship quality as well as its organization (see e.g., Epskamp et al., 2018). Instead of modelling romantic-relationship quality as a reflective or formative latent variable, in which the construct is one common cause of its indicators, relationship quality could perhaps better be conceptualized as clusters of variables with causal interactions (see Fried, 2017b; Schmittmann et al., 2013).

In formative models (where the composite variable summarizes common variation between items), the latent construct cannot be assumed stable across studies, and the interpretations of these different studies' results may be problematic (e.g., interpretational confounding; Howell et al., 2007). In addition, relations between items are modelled as noise. Such modelling ignores the structure of item connections, even the relationships between items may contain information that is relevant for – in this case – understanding romantic-relationship quality.

For instance, some items that could be central to a network (e.g., “Do you love your partner”) may not be easily targeted for relationship therapy. If another item is strongly related to loving one's partner (e.g., “Do you feel listened to by your partner?”), researchers can more easily formulate formal theoretical models to be tested for interventions. Of course, if such theoretical formal models are tested in randomized clinical trials – and are successful - there is a quicker path to successful relationship therapy.

Reflective models (which are modelled as co-occurring because they result from a latent variable) may not entirely account for the correlation between the items either, because of a variety of assumptions that are often unmet (Fried, 2020). For example, it is not uncommon for items to either load on several factors simultaneously or to contain correlated residuals, which leads to the local independence assumption to be unmet (i.e., items are uncorrelated after conditioning on the latent variable; Epskamp et al., 2017). In addition, items with low correlation with others or high cross-loadings are typically deleted even though they may capture essential aspects of the construct (e.g., see McWilliams & Fried, 2018 for a relevant example on adult attachment relationships). McWilliams and Fried (2018), for instance, modelled existing self-report attachment data. Their initial results were perhaps not so surprising: items related to

anxiety, seeking support, and discomfort disclosing feelings clustered with other items. However, and crucially, a trust-related item (typically omitted in attachment questionnaires) bridged the clusters. Altogether, we think that neither the formative nor the reflective approach can fully capture the complexity of romantic-relationship quality.

In recent years, an alternative psychometric approach has emerged that has started addressing the limitations of either approach: Psychological networks. In network models, correlations among self-reported items result from causal interactions between the items themselves (Borsboom, 2008; Borsboom et al., 2011; Cramer et al., 2010; Schmittmann et al., 2013). In psychological networks, items are nodes that are connected by edges representing interactions. To specify the statistical relationship among items, the Gaussian graphical model, a network model for multivariate normal data, has been formalized in recent years. In Gaussian graphical models, edges are understood as partial correlation coefficients between two nodes (e.g., two items) after controlling for all other nodes (also known as conditional dependence relationship, Epskamp & Fried, 2018).

### **Study 2: Conceptualizing Romantic-Relationship Quality from a Network Perspective**

Romantic-relationship quality has been conceptualized through many indicators, resulting in a wide variety of heterogeneous scales available in the literature (see our Studies 1a and b). Network models therefore seem an appropriate alternative to rethink relationship qualities' structure and organization. Instead of modelling romantic-relationship quality or its components as formative or reflective models, network models conceptualize relationship quality as clusters of variables interacting.



While network analysis and factor models roughly provide the same information when trying to retrieve a factor structure when the underlying model is a factor model, loadings between the two approaches differ enough when this is not the case. Cross-loadings between factors, in particular, are more restricted in factor models (as opposed to network models). As a result, if one interprets factor models, the latent variable causes the relationship between variables, whereas for network models, the emergent cause is the relationship between variables (Christensen & Golino, 2021a). Compared to more commonly used statistical approaches, such as exploratory factor analysis, the network approach therefore has at least three advantages in understanding romantic-relationship quality better.

First, network analysis allows for a far more detailed and systematic analysis into the structure of romantic-relationship quality by considering associations between all pairs of indicators or other elements of the phenomenon in question. To do so, Golino and Epskamp (2017) have developed Exploratory Graph Analysis for the scale construction and validation process not only to identify the dimensional structure of construct but also to identify which items belong to which dimension. Exploratory Graph Analysis is based on estimating a network model followed by the detection of different communities. Exploratory Graph Analysis tends to have comparable or even greater accuracy in estimating different communities than comparable methods like factor or parallel analysis (Golino et al., 2020) and thus provides great promise for the assessment of dimensions (see also Christensen et al., 201, 2020, for example applications).

Second, network approaches have clear clinical implications. Several centrality measures, such as node strength or expected influence, can help us identify the most interconnected nodes that might be vital for the functioning of a romantic relationship. In a longitudinal study on depressive disorder, for instance, having higher levels of centrality at baseline was more

predictive of the development of a depressive disorder than having lower levels of centrality at baseline (Boschloo et al., 2016). Developing network models and understanding the centrality of romantic-relationship quality may thus help us identify which couples are potentially at risk for relationship dissolution and the therapist may be able to identify potential intervention mechanisms for positive change. Third, network analysis can provide novel insights into comorbid relationship dysfunction and offer guidance for developing more targeted interventions for individuals experimenting with couples in distress.

Few studies have explored associations between observable indicators of romantic-relationship quality using network analysis. An unpublished study by Li et al. (2021) identified central items and concepts in the marital satisfaction subscale of the ENRICH questionnaire. Among the ten items of the network, the authors found that emotional expressiveness, communication, and understanding were central in the network, potentially suggesting that affective communication and responsiveness may play a key role in diagnostics and interventions. However, the study focused on one specific group (older Chinese individuals), while the assessment was focused on a single instrument (the marital satisfaction subscale of the ENRICH questionnaire). Another study explored the associations between variables measuring sexual and relationship satisfaction and related constructs (Nickull et al., 2022). Associations between most deficiencies that couples (and measurement instruments) often present as a focus of concern, such as deficits in closeness, shared values, trust, joy, or love, have not yet been considered.

### **Study 2 Research Overview**

The main goal of Study 2 was to provide an alternative perspective for the conceptualization and assessment of romantic-relationship quality. Using network analysis, we

aimed to identify the complex association between indicators and identified central items and domains potentially relevant for diagnosis and intervention. First, we estimated a partial correlation network of commonly used items with cross-sectional data from French-speaking participants currently in a romantic relationship. Second, we applied a community detection algorithm to identify the number of communities formed in the network. Third, we determined the most interconnected nodes and identified nodes that acted as bridges between communities. We performed checks for accuracy and stability of the network mainly via bootstrap procedures. To determine key associations among items and between item communities, our approach was exploratory; thus, no hypotheses were generated about specific items that would emerge as central nodes and connectors within the network. Nevertheless, we did assume that global relationship satisfaction/happiness may play a key role, as it has often been used as a global measure of romantic-relationship quality.

## **Methods**

### **Item Selection, Participants, and Procedure**

*Item Selection.* We extracted items that we found in Study 1a, generating an initial set of 50 items by including two items per domain to limit the length of the whole questionnaire. We selected and adapted items from existing scales based on the following criteria: The item has 1) the highest interpretability, 2) the lowest ambiguity, 3) without jargon, 4) not double-barrelled, 5) not mixed positively and negatively, and 6) not too long. Blandine Ribotta and Miguel Silan conducted the selection and agreed on the final inclusion of items. The items were forward-translated from English to French by Blandine Ribotta.

The final questionnaire thus included 50 declarative statements with 7-point Likert response format (from 1 = “Strongly Disagree” to 7 = “Strongly Agree”) assessing participants’

behaviors, beliefs, and feelings toward their relationship and their partner. Seven items were negatively stated but they weren't reverse-scored in order to differentiate positive from negative attitudes in the network. Two child-related items were displayed only to participants with children (and thus only included in a subset of the network analyses). The full survey can be found at <https://osf.io/j42hm>.

**Power Analysis.** To date, there is still limited available evidence to guide a priori power analyses. Epskamp and Fried (2018) suggested that the number of observations must exceed the number of parameters. A priori, we therefore intended to collect  $50 \text{ (nodes)} \times 49/2 = 1225$  participants. However, we were able to collect full data from “only” 587 participants. We therefore adjusted – before any data analysis – the number of nodes to be analyzed so that we only retained 33 nodes (i.e., questions), as that reduced the parameters (edges) to 528 ( $33 \times 32/2$ ). After discussion, we agreed to keep the items that we felt were most important according to common theoretical conceptions, while keeping at least one item per category (except for the Family/Parenting category). The item selection process is available on OSF (<https://osf.io/k5ejt>). We provide the descriptions and node abbreviation for the 33 items included in the analyses in Table 6.

**Participant exclusions and demographics.** As we were interested in participants who were seriously involved in relatively stable romantic relationships, we excluded participants who had been in a relationship for less than six months ( $N = 8$ ) and participants who responded to being casually dating ( $N = 17$ ), divorced ( $N = 8$ ), not seeing anyone ( $N = 7$ ), or others ( $N = 12$ ). In addition, 96 participants were excluded for not answering at least one romantic-relationship quality item. Altogether, we analyzed the data of 587 French-speaking participants ( $M_{age} = 32.4$ ,  $SD_{age} = 12.9$ ; 82.79% women, 15.33% men, 1.53% non-binary/third gender, <1% unidentified). A

majority reported being seriously dating (i.e., 70.19% were not engaged or married) and being in their current relationship for more than three years (58.26%). Finally, a vast majority of the participants reported living on mainland France (89%; see Appendix A). Other characteristics of the sample are detailed in Table 5.

***Procedure.*** We collected data via the survey-platform Qualtrics from March-May 2022. The average time to complete the questionnaire was ~20 minutes. Data were collected from French-speaking (mostly psychology) students (in lab cubicles; rewarded with course credit) and community samples (at home). Student participants were recruited through an advertisement on the lab website; community samples were recruited through multiple websites and Facebook posts. Participants were first asked for their consent and some personal, relationship, and other demographic information. We asked participants for their age, their gender (female, male, non-binary/third gender), the city or country where they live, their relation status (casually dating, seriously dating, engaged, married, divorced, not seeing anyone, or other), the relation duration (6 months to less than 1 year, 1 year to less than 3 years, 3 years to less than 5 years, 5 years to less than 10 years, 10 years or more), their sexual orientation (heterosexual, homosexual, bisexual, asexual, other), and if they have children (and if so, whether they had them with their current partner). Participants also indicated the extent to which their current relationship was monogamous and the extent to which their ideal relationship would be monogamous (5-point scales, ranging from “Entirely monogamous” to “Entirely polyamorous”). Participants then completed the set of romantic-relationship quality items, answering to what extent they agree with a list of 50 declarative statements (from “Strongly Disagree” to “Strongly Agree”; for the full list in French and English, see Appendix B), and the Experiences in Close Relationships

Scale (ECR-R; Fraley et al., 2011).<sup>5</sup> The order of the items within each scale and the order of the scales were randomized.

### **Network Analysis Approach**

Our network analysis procedure consisted of the following steps: First, in Step 1, we estimated a 33-node network, computed node centrality expected influence, and checked the accuracy and stability of these estimates. The second phase of Step 1 was to use Exploratory Graph Analysis to identify communities of nodes, to identify the dimensionality, and to evaluate the stability of the dimensions and items. Once we had removed items that reduced our solution's stability, we moved to Step 2 and again checked the stability of the solution after which we again identified communities of nodes and the dimensionality and evaluated the stability of the dimensions and items. Once the dimensions had good structural consistencies, we computed bridge expected influence. We conducted all analyses using RStudio (RStudio Team, 2020; v.4.0.2) on May 11th, 2022. All data, cleaning procedures, R code, and additional packages to carry out the analysis are available on the Open Science Framework for reproduction and replication purposes: <https://osf.io/6y723>.

*Network Estimation and Visualization.* To estimate and visualize the network, we used the estimateNetwork function from the bootnet package (Epskamp et al., 2018; v.1.5) and the plot function from the qgraph package (Epskamp et al., 2012; v.1.9.1). We estimated a Gaussian

---

<sup>5</sup> We included a measure of romantic attachment style to provide participants with feedback regarding their scores on the anxiety and avoidance dimensions. These scores are interpreted in terms of general attachment style (secure/preoccupied/dismissing/fearful-avoidant) by combining the scores from each of the two subscales. Copies of data could be provided upon request (the template can be found at <https://osf.io/d2njw>). Details of the attachment scale and the feedback provided can be found at the end of the full survey (available at <https://osf.io/j42hm>).

graphical model using graphical LASSO based on the extended Bayesian information criterion (Epskamp & Fried, 2018; Foygel & Drton, 2010; Friedman et al., 2008). This procedure accounted for potential spurious relations by addressing the risk of false positive connections by shrinking all edges and setting small edges to zero.

To balance the sensitivity and specificity of identifying true edges, we set the EBIC hyper-tuning parameter (specifying the sparsity level) to 0.5 (Foygel & Drton, 2010). As our data were polytomous and non-normal, we estimated a Spearman correlation matrix as input. As suggested by Epskamp and Fried (2018), we opted for Spearman correlations instead of polychoric correlations because the highly skewed scores of the items can lead to biased polychoric correlations. Of all 33 variables included in the network analysis, 8% of data were missing. Here, we assume that these data were missing at random (Rubin, 1976). Missing data were managed using the complete pairwise observations via the `cor` function.

Our network layout was based on the Fruchterman–Reingold algorithm. This algorithm places nodes with stronger connections near the center of the network and nodes with weaker connections near the periphery (Fruchterman & Reingold, 1991). We checked the edge-weight accuracy with the `bootnet` function. We estimated bootstrapped confidence intervals using nonparametric bootstrapping based on 1,500 bootstrap samples. Wide confidence intervals imply caution in interpretation, especially when interpreting the importance of edges or the presence of weaker edges. The wider the bootstrapped confidence interval is for one edge, the less confidence we can attach to the estimate, and the more careful our inferences should be.

Finally, we computed bootstrapped difference tests to investigate if edge weights significantly differ from one another (Epskamp et al., 2018). The test is based on calculating the

difference between the two bootstrap values (i.e., for the two edge weights) and subsequently testing if the bootstrapped confidence interval around this difference estimate includes zero.

***Community Detection and Dimensionality.*** To identify whether the items formed distinct communities within the network, we used the **Exploratory Graph Analysis** function from the EGAnet package (Golino & Christensen, 2020; v.1.0.1). **Exploratory Graph Analysis** first applies a network estimation method followed by a community detection algorithm (Fortunato, 2010). We estimated a Gaussian graphical model via the EBICglasso algorithm (Foygel & Drton 2010). We then applied the Louvain community detection algorithm because Christensen et al. (under review) found that the Louvain algorithm performs comparably or better than the Walktrap or spinglass algorithms (which have typically been adopted in the network literature).

We scrutinized the dimensionality of the structure by examining network loadings and structural consistency with the bootEGA function (Christensen & Golino, 2021a). Network loadings (also called standardized node strength) can be interpreted in the same way as an exploratory factor analysis load matrix and describe the extent to which a node contributes to a dimension in the network (net.loads function; Christensen & Golino, 2021b). However, because these loadings represent partial correlations, they are much smaller than the factor loadings (Christensen et al., 2021b). No consensus has been yet established to interpret these loadings; however, Christensen et al. (2020) have recommended effect size guidelines for network loadings to be 0.10, 0.30, and 0.50 for small, moderate, and large effects, respectively. Furthermore, we estimated the structural consistency of each dimension by using the bootEGA procedure with 1,500 iterations.



**Centrality Indices.** After network estimation, we computed node expected influence to quantify the cumulative importance of nodes within a network (Robinaugh et al., 2016); we used the `centralityPlot` and `centralityTable` functions in the `qgraph` package. Node expected influence is a measure of centrality that takes the sign of edge weights into account and is defined as the sum of the value of all edges connecting to a specific node. The higher values reflect greater node influence within the network.

After identifying communities, we computed bridge expected influence to examine if any items acted as bridges through the `networktools` package (Jones, 2020; v.1.4.0). This bridging metric focuses solely on connections between communities and determine which nodes were influential in bridging item communities. Both centrality estimates were presented as standardized z scores, with higher scores (positive or negative) reflecting a higher degree of importance within a network. We investigated the stability of the expected influence estimates for both indices by using case-drop bootstrapping based on 1,500 bootstrap samples. Case-dropping bootstrap assesses the degree to which centrality estimates are subject to sampling error by correlating the centrality indices from the whole sample with centrality indices obtained through estimating networks on sample subsets.

We also computed the correlation stability coefficient (CS coefficient; Epskamp et al., 2018). The larger the CS coefficient, the more reliable the interpretation of the centrality order estimates. We computed this coefficient using the `corStability` function from `qgraph` (the `corStability` function should not be below 0.25 and preferably above 0.5; Epskamp et al., 2018). Finally, we computed bootstrapped difference tests to investigate whether centrality indices differ significantly from one another (Epskamp et al., 2018). The test is based on calculating the difference between the two bootstrap values (i.e., the two centrality indices).

## Results

The results of our stability and precision analyses are available in the appendices for the sake of brevity. All the results of the network analyses, descriptive statistics of all variables measured as well as the frequency of responses to ordinal variables are available on our OSF page: <https://osf.io/jvr58>. Our final solution converged to four dimensions: 1) shared life philosophies, joint decision-making, and trust, 2) one's own love and the emotional and physical connection with the partner, 3) agreement or disagreement in the relationship and the ability to cope with disagreement, and 4) perceptions of the partner pertaining to commitment, love, and the feeling of being needed.

“Things are often going well between us”, “My relationship with my partner makes me happy”, and “I don't get the love and affection I want from my partner” were most central to the network, as they exhibited the highest connectivity with other items or with various clusters of items. In what follows, we detail how we arrived at these conclusions.

### Step 1: Preliminary Network Analysis of All 33 Items

The network visualization is presented in Figure 5A. Overall, the network contained 240 non-zero edges (46%) among 528 possible edges, of which a majority were positive (78%). Six positive edges and two negative edges have an absolute weight greater than .20. These edges are: “I often feel angry or resentful toward my partner” and “We often quarrel” (weight = 0.28), “When there is a difference of opinion, we try to talk it out rather than fight” and “I am very happy with the way we talk and communicate” (weight = 0.26), “This relationship with my partner satisfies my need” and “My relationship with my partner makes me happy” (weight = 0.23), “I don't get the love and affection I want from my partner” and “My partner is committed to our relationship”) (weight = -0.22), “My partner and I often engage in outside interests

together” and “My partner and I often have fun together” (weight = 0.22), “I often want to hug and cuddle my partner” and “I am happy with our sexual relationship” (weight = 0.21), “I don't get the love and affection I want from my partner” and “This relationship with my partner satisfies my need” (weight = - 0.20), “I love my partner” and “I am committed to our relationship” (weight = 0.20).

The results of accuracy analysis indicated that many of the strongest edges had 95% confidence intervals for the edge weights did not overlap, suggesting that they were significantly stronger (Appendix C; Figure C1). Bootstrapped edge-weight difference tests supported this interpretation (Appendix C; Figure C2). The expected influence estimates are shown in Figure 5B. The node “I don't get the love and affection I want from my partner” was particularly highly and connected negatively with other nodes in the network (z-score = -3.00). Five other nodes exhibit a relatively high z-score: The nodes “I have a strong connection with my partner” (z-scores = 1.54) and “My relationship with my partner makes me happy” (z-score = 1.53) were highly positively connected with other nodes. The nodes “We often discuss or consider divorce, separation, or terminating our relationship” (z-scores = -1.58), “My partner often tries to control or influence my life” (z-score = -1.54) and “We have large differences in our views towards life” (z-score = -1.45) were highly negatively connected with other nodes. Stability of node centrality indices showed that expected influence estimates were stable, with a centrality stability coefficient of 0.67, indicating that 67% of the data could be dropped to retain with 95% certainty a correlation of 0.7 with the original dataset (Appendix C; Figure C3). Bootstrapped difference tests at a significance level of 0.05 indicated that most node centralities can be shown to significantly differ from each other (Appendix C; Figure C4).

The Exploratory Graph Analysis revealed a four-cluster solution, shown in Figure 5A. Before any interpretation, we checked the stability of the Exploratory Graph Analysis results (see Appendix D). Four dimensions were found 80.30% of the time while five dimensions were found 16.10% of the time. These results suggest that the four-dimension solution might be unstable. Three of the four dimensions showed a low structural consistency (Dimension 1 = 0.22; Dimension 2 = 0.29; Dimension 3 = 0.05; Dimension 4 = 0.94). Six items of these three dimensions had low stability (i.e., were at below of the range of 0.60; see Figure 5C and Appendix D; Table D4), that is, these items do not often replicate in their empirical Exploratory Graph Analysis defined dimension. Some items had cross loadings worth consideration ( $\geq 0.10$ ) or have a cross-loading approaching the same threshold (Appendix D; Table D5). These results suggest that some items are multidimensional and are decreasing the structural consistency of the dimensions. We decided to remove these items, as this may improve the stability of the dimensions and better structural consistency of the scale.

## **Step 2: Item Reduction and Final Stability of Four Dimensions of Romantic Relationship Quality**

Based on item stability statistics, centrality indices, and theoretical considerations, we identified and removed nine unstable items to reach four stable dimensions. The removed items were: “It is easy for me to forgive my partner”, “Our personalities are rather matched”, “I'm happy with the way my partner takes care of things at home”, “My partner encourages me to do the things I think are important” (initially from Dimension 1); “This relationship with my partner satisfies my need” (Dimension 2); “I often

depend on my partner”, “My partner often tries to control or influence my life”, “My relationship with my partner is very stable”, “We often discuss or consider divorce, separation, or terminating our relationship” (Dimension 3). The unstable item “My relationship with my partner makes me happy” was not removed because it is one of the most central nodes of the network. The 24-node network and the four estimated communities are shown in Figure 6A. Structural consistency of the four dimensions were sufficient (Appendix F; Table F) and item stability statistics were above 0.81 (Figure 6B). We provide all results of the dimensionality stability analyses in Appendix F.

Before identifying what the new estimated dimensions represent, we analyzed the newly estimated network and its accuracy. After removing the nine items, the network contained 24 nodes and 149 non-zero edges (49%) among 276 possible edges. The highest edges were quite like the 33-node network, with the highest one being identical: “I often feel angry or resentful toward my partner” and “We often quarrel” (weight = 0.29).

Bootstrapped confidence intervals and difference tests between all pairs of edge weights indicated great accuracy of the strongest edges (Table E1 and E2). The two most central nodes remain the same (Figure 6C), with “I don't get the love and affection I want from my partner” (z-score = -2.86) and “My relationship with my partner makes me happy” (z-score = 1.81). The expected influence estimates were sufficient, with a centrality stability coefficient of 0.75 (Figures E3 and E4; All results of the accuracy and stability analyses of the estimated network are provided in the Appendix E).

We then interpreted the different dimensions. Dimension 1 contained nine items reflecting one's own love and the emotional and physical connection with the partner ("I love my partner", "I have a strong connection with my partner", "There are things I do only with my partner", "I often want to hug and cuddle my partner", "I am committed to our relationship", "My relationship with my partner makes me happy", "I am happy with our sexual relationship", "My partner and I often engage in outside interests together", "My partner and I often have fun together"). Dimension 2 contained six items reflecting shared life philosophies, joint decision-making, and trust ("My partner and I agree to make important decisions", "I sometimes worry about his/her behaviors on social occasions", "We have large differences in our views towards life", "My partner and I agree on our dealings with our in-laws", "My partner and I agree regarding friends", "I trust my partner"). Dimension 3 contained five items reflecting agreement or disagreement in the relationship and the ability to cope with disagreement ("When there is a difference of opinion, we try to talk it out rather than fight", "I am very happy with the way we talk and communicate", "Things are often going well between us", "We often quarrel", "I often feel angry or resentful toward my partner"). Dimension 4 contained four items reflecting perceptions of the partner pertaining to commitment, love, and the feeling of being needed ("I don't get the love and affection I want from my partner", "For me, my partner is the perfect romantic partner", "My partner makes me feel needed", "My partner is committed to our relationship"). The loadings for

items on each of their respective communities are in the small range (0.11-0.29, 0.16-0.23, 0.23-0.26, 0.17-0.29, for Dimension 1, 2, 3, and 4 respectively).

The bridge centrality index of expected influence indicated that “Things are often going well between us” and “I don't get the love and affection I want from my partner” exhibited the highest connectivity with the various clusters of items (Figure 6D). These results were confirmed by examining the network cross-loadings, ( $\geq 0.10$ ): “My relationship with my partner makes me happy” with Dimension 3 and 4 (0.15 and 0.15); “My partner is committed to our relationship”, “For me, my partner is the perfect romantic partner”, and “Things are often going well between us” with Dimension 1 (0.10, 0.12, and 0.13). Stability analysis showed that bridge expected influence estimates were stable, with a centrality stability coefficient of 0.75 (Figure E5), and bootstrapped difference tests indicated that the four strongest bridges centralities can be shown to significantly differ from each other (Figure E6).

**Auxiliary analyses.** We performed four additional analyses by adding four additional variables to the 33-items network in separate analyses. We obtained 4 distinct 34-items networks. We found that the anxious attachment score was negatively and strongly correlated with the item “I trust my partner” (-0.19), and was positively and weakly correlated with the items “I often depend on my partner” (0.1), “I don't get the love and affection I want from my partner” (0.08), “I often want to hug and cuddle my partner” (0.08),

"We often discuss or consider divorce, separation, or terminating our relationship" (0.08), and "I sometimes worry about his/her behaviors on social occasions" (0.07). The avoidant attachment score was weakly and negatively correlated with several items such as "I have a strong connection with my partner" (-0.11), "I trust my partner" (-0.09), "There are things I do only with my partner" (-0.09), "I am committed to our relationship" (-0.09), and "I often depend on my partner" (-0.06).

The "ideal" level of monogamy was weakly and negatively correlated with "There are things I do only with my partner" (-0.12) and positively correlated with the items "We often discuss or consider divorce, separation, or terminating our relationship" (0.08), and weakly correlated with "I love my partner" (-0.05). The level of "current" monogamy was weakly and negatively correlated with the item "My relationship with my partner is very stable" (-0.07).

### **Discussion Study 2**

As the literature on the assessment on romantic-relationship quality exhibited considerable heterogeneity (Study 1a) and showed limited evidence of internal structure validity (Study 1b), we decided to conduct a new validation study with 33 items from the existing literature that we had, a priori, defined as being part of 25 different categories. After removing 9 items due to stability issues, we found a solution consisting of four main dimensions of romantic-relationship quality: 1) one's own love and the emotional and physical connection with the partner, 2) shared life philosophies, joint decision-making, and trust, 3) agreement or disagreement in the relationship and the ability to cope with disagreement, and 4) perceptions of



the partner pertaining to commitment, love, and the feeling of being needed. Central to the entire network of romantic-relationship quality are the items “Things are often going well between us”, “My relationship with my partner makes me happy”, and “I don't get the love and affection I want from my partner”.

### **What the network of romantic-relationship quality means**

The mutual interaction between the different indicators in the network yielded highly interconnected communities. Feelings of love, connectedness, commitment, and exclusiveness were highly interconnected with each other and were associated with the desire to be hugged, which in turn was highly related to sexual satisfaction and having fun. These strong links indicate that these indicators are all closely related concepts and that changing one affects another node in the network. Among all the indicators, having fun and feeling in love were strongly associated with relationship happiness, whereas sexual satisfaction was relatively very weakly connected to relationship happiness (0.02 and 0.07 for the 33- and 24-node network). The low partial correlation between the two indicators is explained to a considerable extent by another indicator, as sexual satisfaction and relationship happiness were both strongly linked to need fulfilment. This result is supported by the edge-weight difference between the two networks, where the degree of association drops drastically when this indicator is removed from the network. This finding suggests that sex could be part of the needs that are met by the partner, which could, in turn, increase happiness in the relationship.

The present study makes a unique contribution to the large body of literature on romantic-relationship quality and provides a novel insight into the dynamics of the construct of romantic relationship quality. Romantic relationships should not be characterized by “simple” latent variables, but rather, as a network system. From our analyses, we infer there to be at least a

reciprocal relation between different nodes, such as the question “I love my partner” and “I am committed to our relationship”.<sup>6</sup>

### **Recommendations for which items to use in assessing romantic-relationship quality**

Ideally, researchers or therapists use the 24 items that we identified after item reduction. However, they can choose to stick to only three items if they are short on time. The questions “Things are often going well between us”, “My relationship with my partner makes me happy”, and “I don't get the love and affection I want from my partner” were the most central in the network and are thus the first three items that should be targeted. The questions “Things are often going well between us” and “I don't get the love and affection I want from my partner” exhibited the highest bridge centrality estimates, suggesting that these indicators had a greater effect on indicators *outside* their community and should therefore definitely be part of the three items that are asked. Remarkably, other indicators were not as central as one might expect, such as the question “I am happy with our sexual relationship” or the question “Our personalities are rather matched” and can thus be dropped from one's assessment.

### **Insights for therapy**

Our findings may also provide clinical insights. First, we can modify (i.e., activate or inhibit) central nodes, which in turn, may help to trigger, develop, and maintain a good relationship functioning. It may be hard for couples (or therapists helping couples) to directly improve on matters like “Things are often going well between us”, “My relationship with my

---

<sup>6</sup> Had one predicted the final list of items of our scale based on their frequency of occurrence in past instruments, one would have predicted an item from affection/love and emotional closeness/intimacy to be central to the network. Instead, this was affection/love and global happiness/satisfaction. Similarly, one would have (erroneously) included expectations/needs met, emotional support, conflict resolution, power/control, and stability to be part of the scale. Conversely, we would have missed mutual dependence”, “exclusiveness/monogamy”, “physical affection”, and “agreement about proper conduct” to be part of the final scale.

partner makes me happy”, and “I don't get the love and affection I want from my partner”. If one wants to increase happiness in a relationship, how would one even go about that?

Instead, we recommend targeting some of the aspects that are directly and strongly linked to these nodes. For instance, “I don't get the love and affection I want from my partner” is strongly linked to “My partner makes me feel needed”. Therapists could attempt to understand what makes the partner feel more needed to increase feelings of love. Similarly, “Things are often going well between us” is strongly linked to “We often quarrel”, “I am very happy with the way we talk and communicate”, and “My partner and I agree to make important decisions”, implying that the way to deal with disagreements in relationships can be targeted to improve partners' sense of well-being in the relationship. Finally, “My relationship with my partner makes me happy” is strongly linked to “My partner and I often have fun together” and “My partner and I often engage in outside interests together” implying that increasing shared activities may improve the sense of exclusivity and commitment to the partner, which may increase the sense of connectedness and satisfaction, all with, separately or collectively, may feed then back to raise positive feelings or reduce negative feelings on other dimensions. Of course, we are making causal inferences where none may be merited given the cross-sectional design; rigorous randomized control trials are needed to establish whether such causal links exist. In the absence of this, however, this is the readiest the evidence available (IJzerman et al., 2020).

Our cross-sectional design has other limitations. While relationship quality is typically conceptualized as a within-person evaluation of the romantic relationship, or the quality of the dyadic interaction in itself – what is commonly done is a between-person factor analysis. This mismatch between conceptualization and measurement model is worrisome, as validity cannot be determined in case of a mismatch. Items assessing romantic-relationship quality often measure

multiple categories and scales are likely not unidimensional. Most scales are therefore likely to violate simple structure in a factor model, which, if one takes understanding the matter of relationship quality seriously, for the different categories involved would invariably lead to poor model fit. While relationship quality can be characterized as a within-person model, we believe that the best way to model relationship quality is through a within-person coupled with a within-dyad approach *over time* (Hamaker, 2012; Hamaker & Wichers, 2017).

### **Constraints on generality (Simons et al., 2017)**

There are other important limitations that we have to outline. Although our sample had quite some variation in relationship duration, age, and many were above student age, our sample consisted mostly of heterosexual French women that preferred monogamy in their relationship. We find it unlikely that our results will replicate in samples with other nationalities, with other sexual orientations, other genders, and (for any kind of) less-monogamous individuals. Further, our sample was restricted in its range as it consisted of individuals who reported to be relatively high in romantic-relationship quality (with means of 5.06 for “Things are often going well between us”, 5.86 for “My relationship with my partner makes me happy”, and 2.60 “I don't get the love and affection I want from my partner”). We therefore doubt that, for instance, couples in therapy would have the same network of romantic-relationship quality. But in the absence of more empirical work, what is a therapist to do? We still recommend using our three items with the proposed interventions, with a healthy dose of skepticism based on the therapist's experience.

### **Other current challenges**

Some scales synonymize relationship quality with relationship satisfaction (CSI, KMSS) or otherwise with “adjustment” (DAS, MAT, RDAS, RMAT) or the accommodation of the partner to one another. However, many either do not define, or have no clear conceptualization of

relationship quality -- instead presuming ad-hoc constructed and empirically derived factors (from a between-person factor analytic model) constitute the important components of relationship quality (e.g., compatibility, affection, sexual interest, and so forth).

This causes two massive challenges for the measurement of relationship quality. First, relationship quality is a fuzzy construct; what relationship quality is constituted of is likely to differ from couple to couple and from culture to culture. There are bound to be shared commonalities (e.g., love, trust, reliability, and so forth) but these are likely non-sufficient or non-necessary for any particular couple. It's unlikely for fuzzy constructs, including relationship quality, to have necessary and sufficient conditions (Silan et al., 2022). Rather than prematurely defining what relationship quality is, an exploratory study that explicates folk logic through qualitative methods would provide valuable information to this *constitutivity* issue. This includes information about at which level (at the couple, country, or regional culture level?) a standardized scale is useful, and which items should be included in a potentially cross-culturally generalizable relationship quality scale.

Another challenge for relationship quality is that it's commonly defined as a within-person subjective evaluation of the relationship. Yet romantic relationship quality has been commonly developed and validated are through between-person factor analytic models. This mismatch leaves a concerning gap on our understanding of relationship quality. Thus, dynamic modelling techniques such as longitudinal network models (e.g., Snijders, 2005) or complexity methods provide fruitful avenues forward. However, measurement models that deal both with temporal dynamics and fuzzy constitutivity are still yet to be developed.

**How to solve our constraints on generality and other future directions**

First, one can take a big-team science approach to the romantic-relationship literature (Forscher et al., 2022), first making a qualitative inventory of what romantic relationship quality is across regions, then developing measures together with community members and validating them thereafter across the world (Dujols et al., 2022). This work is currently in progress (Silan et al., 2023). A second interesting alternative can be derived from the meta-analysis literature, where researchers have started moving to “community-augmented meta-analysis” (see, for instance, Sparacio et al., 2023a, b; IJzerman et al., 2022; Tsuji et al., 2014). We think that relationship science, and psychology more generally, would be served by comparable initiatives for measurement. Research waste would be reduced if grant agencies invest into systematic measurement databases on which researchers and practitioners can draw when using and developing measures. These measurement databases can take two forms. First, researchers can pool responses to questions, and measures can be developed and validated through these large databases. Second, measurement databases can report on item overlap and general validity evidence across measures, so that researchers know which measures to utilize.

## **Conclusion**

Despite all the limitations we have outlined, our project provides a generative launching point to a better conceptualization of relationship quality. Study 1A and B show the various measurement issues that need to be overcome, from conceptualization to validation. Study 2 provides initial insights into between-person associations and serves as a proof-of-concept for how to measure romantic relationship quality and for alternative modeling techniques to be used for relationship quality.



*Author Contributions:* BR wrote the first draft of the manuscript for her master's thesis and OD, MAS, and HIJ provided critical comments throughout the thesis project. BR conducted the first analyses and CvL conducted code review at the end of the project. After we found a number of shortcomings and the code was not available anymore for Study 1a, BP rewrote the analysis script for Study 1a and corrected a number of mistakes and added where there were omissions. BR conducted the search and the rating of the item-content overlap; OD, MAS, and HIJ did a partial double-check of the search and ratings. HIJ rewrote the manuscript several times.

*Conflict of Interest:* Miguel Silan is working on an instrument to assess relationship quality that may be sold to cover the investment into the development costs that pays for his PhD project (after which the instrument will become free). Hans IJzerman and Miguel Silan both work for a behavioral science startup (Annecy Behavioral Science Lab) that focuses on reducing social isolation.



### References

- Ahlborg, T., Lilleengen, A. M., Lönnfjord, V., & Petersen, C. (2009). Quality of dyadic relationship in Swedish men and women living in long-term relationships and in couples in family counselling—Introduction of a new self-report measure, QDR36. *Nordic Psychology, 61*(3), 23-46. <https://doi.org/10.1027/1901-2276.61.3.23>
- American Educational Research Association. (2014). *Standards for educational and psychological testing*. American Educational Research Association American Psychological Association National Council on Measurement in Education.
- Arrindell, W. A., Boelens, W., & Lambert, H. (1983). On the psychometric properties of the Maudsley Marital Questionnaire (MMQ): Evaluation of self-ratings in distressed and ‘normal’ volunteer couples based on the Dutch version. *Personality and Individual Differences, 4*(3), 293-306. [https://doi.org/10.1016/0191-8869\(83\)90151-4](https://doi.org/10.1016/0191-8869(83)90151-4)
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications.
- Belous, C. K., & Wampler, R. S. (2016). Development of the gay and lesbian relationship satisfaction scale. *Journal of Marital and Family Therapy, 42*(3), 451-465. <https://doi.org/10.1111/jmft.12158>
- Borsboom, D. (2008). Latent variable theory. *Measurement, 6*(1 & 2), 25-53. <https://doi.org/10.1080/15366360802035497>
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review, 111*(4), 1061. DOI: 10.1037/0033-295X.111.4.1061

- Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, 6(11), e27407. DOI: 10.1371/journal.pone.0027407
- Boschloo, L., Borkulo, C. D. van, Borsboom, D., & Schoevers, R. A. (2016). A prospective study on how symptoms in a network predict the onset of depression. *Psychotherapy and Psychosomatics*, 85(3), 183–184. <https://doi.org/10.1159/000442001>
- Blum, J. S., & Mehrabian, A. (1999). Personality and temperament correlates of marital satisfaction. *Journal of Personality*, 67(1), 93-125. <https://doi.org/10.1111/1467-6494.00049>
- Braiker, H. B., & Kelley, H. H. (1979). Conflict in the development of close relationships. In R. L. Burgess & T. L. Huston (Eds.), *Social exchange in developing relationships*. New York, NY: Academic.
- Busby, D. M., Christensen, C., Crane, D. R., & Larson, J. H. (1995). A revision of the Dyadic Adjustment Scale for use with distressed and nondistressed couples: Construct hierarchy and multidimensional scales. *Journal of Marital and Family Therapy*, 21(3), 289-308. <https://doi.org/10.1111/j.1752-0606.1995.tb00163.x>
- Cheng, Z. H., Tan, L. X., Yang, Y., Lin, X. H., Zhou, D., Jiang, X., Su T. H., Zhao Y. & Yuchi, X. L. (2010). The Chinese marital quality inventory: development, reliability and validity. In Lawrence T. Lam, Shi-Jie Zhou, & Guo-Bin Wn (Eds), *Psychological and Health-related Assessment Tools Developed in China*, (pp. 61–70) Bentham Science Publishers Ltd. <https://doi.org/10.2174/978160805186111001010061>

- Christensen, A. P., & Golino, H. (2021a). On the equivalency of factor and network loadings. *Behavior Research Methods*, 53(4), 1563–1580. <https://doi.org/10.3758/s13428-020-01500-6>
- Christensen, A. P., & Golino, H. (2021b). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. *Psych*, 3(3), 479–500. <https://doi.org/10.3390/psych3030032>
- Christensen, A. P., Golino, H., & Silvia, P. J. (2020). A psychometric network perspective on the validity and validation of personality trait questionnaires. *European Journal of Personality*, 34(6), 1095–1108. <https://doi.org/10.1002/per.2265>
- Christensen, A. P., Gross, G. M., Golino, H. F., Silvia, P. J., & Kwapil, T. R. (2019). Exploratory graph analysis of the multidimensional schizotypy scale. *Schizophrenia Research*, 206, 43–51. <https://doi.org/10.1016/j.schres.2018.12.018>
- Cramer, A. O., Waldorp, L. J., Van Der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, 33(2-3), 137-150. DOI: 10.1017/S0140525X09991567
- Davis, K. E. (2001). The relationship rating form (RRF): A measure of characteristics of romantic relationships and friendships. In J. Touliatos, B. Perlmutter, & G. Holden (Eds.) *Handbook of family measurement techniques*. (2nd ed., pp. 195-197). Sage Publications.
- Delatorre, M. Z., & Wagner, A. (2020). Marital quality assessment: Reviewing the concept, instruments, and methods. *Marriage & Family Review*, 56(3), 193–216. <https://doi.org/10.1080/01494929.2020.1712300>

- Ducat, W. H., & Zimmer-Gembeck, M. J. (2010). Romantic partner behaviours as social context: Measuring six dimensions of relationships. *Journal of Relationships Research, 1*(1), 1-16. <https://doi.org/10.1375/jrr.1.1.1>
- Dujols, O., Klein, R. A., Zsido, A. N., Szymkow, A., Zickfeld, J., ... IJzerman, H., (2023, January 30). STRAEQ-2 Home. Retrieved from [osf.io/ghbzk](https://osf.io/ghbzk)
- Dunn, T. J., Baguley, T., & Brunsten, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105*(3), 399-412. <https://doi.org/10.1111/bjop.12046>
- Epskamp, S., & Fried, E. I. (2018). A tutorial on regularized partial correlation networks. *Psychological Methods, 23*(4), 617. DOI: 10.1037/met0000167
- Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods, 50*(1), 195-212. <http://dx.doi.org/10.3758/s13428-017-0862-1>
- Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika, 82*, 904-927. <https://doi.org/10.1007/s11336-017-9557-x>
- Epskamp, S., Cramer, A., Waldorp, L., Schmittmann, V., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*. <https://doi.org/10.18637/jss.v048.i04>
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co.

- Faelens, L., Hoorelbeke, K., Fried, E., De Raedt, R., & Koster, E. H. W. (2019). Negative influences of Facebook use through the lens of network analysis. *Computers in Human Behavior, 96*, 13–22. <https://doi.org/10.1016/j.chb.2019.02.002>
- Fellows, K. J., Chiu, H.-Y., Hill, E. J., & Hawkins, A. J. (2016). Work–family conflict and couple relationship quality: A meta-analytic study. *Journal of Family and Economic Issues, 37*(4), 509–518. <https://doi.org/10.1007/s10834-015-9450-7>
- Fincham, F. D., & Bradbury, T. N. (1987). The assessment of marital quality: A reevaluation. *Journal of Marriage and the Family, 49*(4), 797. <https://doi.org/10.2307/351973>
- Fincham, F. D., & Linfield, K. J. (1997). A new look at marital quality: Can spouses feel positive and negative about their marriage?. *Journal of Family Psychology, 11*(4), 489. <https://doi.org/10.1037/0893-3200.11.4.489-502>
- Fincham, F. D., & Rogge, R. (2010). Understanding relationship quality: Theoretical challenges and new tools for assessment. *Journal of Family Theory & Review, 2*(4), 227–242. <https://doi.org/10.1111/j.1756-2589.2010.00059.x>
- Fletcher, G. J. O., Simpson, J. A., & Thomas, G. (2000). The measurement of perceived relationship quality components: A confirmatory factor analytic approach. *Personality and Social Psychology Bulletin, 26*(3), 340–354. <https://doi.org/10.1177/0146167200265007>
- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science, 3*(4), 484-501. <https://doi.org/10.1177/2515245920951747>

- Flake, J. K. (2021). Strengthening the foundation of educational psychology by integrating construct validation into open science reform. *Educational Psychologist*, 56(2), 132-141. <https://doi.org/10.1080/00461520.2021.1898962>
- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, 8(4), 370–378. <https://doi.org/10.1177/1948550617693063>
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 456–465. <https://doi.org/10.1177/2515245920952393>
- Fletcher, G. J., Simpson, J. A., & Thomas, G. (2000). The measurement of perceived relationship quality components: A confirmatory factor analytic approach. *Personality and Social Psychology Bulletin*, 26(3), 340-354. <https://doi.org/10.1177/0146167200265007>
- Forscher, P. S., Wagenmakers, E. J., Coles, N. A., Silan, M. A., Dutra, N., Basnight-Brown, D., & IJzerman, H. (2022). The benefits, barriers, and risks of big-team science. *Perspectives on Psychological Science*, 17456916221082970. <https://doi.org/10.1177/17456916221082970>
- Fowers, B. J., & Owenz, M. B. (2010). A eudaimonic theory of marital quality. *Journal of Family Theory & Review*, 2(4), 334–352. <https://doi.org/10.1111/j.1756-2589.2010.00065.x>
- Foygel, R., & Drton, M. (2010). Extended Bayesian Information Criteria for Gaussian Graphical Models. *arXiv*. <https://doi.org/10.48550/arXiv.1011.6640>
- Fraley, R. C., Heffernan, M. E., Vicary, A. M., & Brumbaugh, C. C. (2011). The experiences in close relationships—Relationship Structures Questionnaire: A method for assessing

- attachment orientations across relationships. *Psychological Assessment*, 23(3), 615. DOI: 10.1037/a0022898
- Fried, E. I. (2017a). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, 208, 191–197. <https://doi.org/10.1016/j.jad.2016.10.019>
- Fried, E. I. (2017b). What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. *Health Psychology Review*, 11(2), 130–134. <https://doi.org/10.1080/17437199.2017.1306718>
- Fried, E. I., & Flake, J. K. (2018). Measurement matters. *APS Observer*, 31(3). <https://www.psychologicalscience.org/observer/measurement-matters>
- Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432–441. <https://doi.org/10.1093/biostatistics/kxm045>
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129–1164. <https://doi.org/10.1002/spe.4380211102>
- Funk, J. L., & Rogge, R. D. (2007). Testing the ruler with item response theory: increasing precision of measurement for relationship satisfaction with the Couples Satisfaction Index. *Journal of Family Psychology*, 21(4), 572. <https://doi.org/10.1037/0893-3200.21.4.572>
- Galovan, A. M., & Schramm, D. G. (2018). Strong relationality and ethical responsiveness: A framework and conceptual model for family science. *Journal of Family Theory & Review*, 10(1), 199-218. <https://doi.org/10.1111/jftr.12238>

- Garthoeffner, J. L., Henry, C. S., & Robinson, L. C. (1993). The Modified Interpersonal Relationship Scale: Reliability and validity. *Psychological Reports*, 73(3), 995–1004. <https://doi.org/10.1177/00332941930733pt141>
- Glenn, N. D. (1990). Quantitative research on marital quality in the 1980s: A critical review. *Journal of Marriage and the Family*, 52(4), 818. <https://doi.org/10.2307/353304>
- Golino, H., & Christensen, A. (2020). eganet: Exploratory Graph Analysis - A framework for dimensionality assessment and reduction using network psychometrics. <https://rdrr.io/cran/EGAnet/>
- Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS one*, 12(6), e0174035. <https://doi.org/10.1371/journal.pone.0174035>
- Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., ... & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, 25(3), 292. <https://doi.org/10.1037/met0000255>
- Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press.
- Hamaker, E. L., & Wichers, M. (2017). No time like the present: Discovering the hidden dynamics in intensive longitudinal data. *Current Directions in Psychological Science*, 26(1), 10-15.
- Hassebrauck, M., & Fehr, B. (2002). Dimensions of relationship quality. *Personal Relationships*, 9(3), 253–270. <https://doi.org/10.1111/1475-6811.00017>



- Hendrick, S. S. (1988). A generic measure of relationship satisfaction. *Journal of Marriage and the Family*, 50(1), 93-98. <https://doi.org/10.2307/352430>
- Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods*, 12(2), 205–218. <https://doi.org/10.1037/1082-989X.12.2.205>
- Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, 3(2), 166–184. <https://doi.org/10.1177/2515245919882903>
- IJzerman, H., Hadi, R., Coles, N. A., Paris, B., Elisa, S., Fritz, W., ... & Ropovik, I. (2021). Social thermoregulation: A meta-analysis. Available via PsyArxiv: <https://psyarxiv.com/fc6yq/>.
- Jones, P. (2020). networktools: Tools for identifying important nodes in networks. <https://cran.r-project.org/web/packages/networktools/>
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112(3), 527–535. <https://doi.org/10.1037/0033-2909.112.3.527>
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>
- Karney, B. R., & Bradbury, T. N. (1995). The longitudinal course of marital quality and stability: A review of theory, methods, and research. *Psychological bulletin*, 118(1), 3. <https://doi.org/10.1037/0033-2909.118.1.3>
- Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Company.
- Kimmel, D., & Van Der Veen, F. (1974). Factors of marital adjustment in Locke's Marital Adjustment Test. *Journal of Marriage and the Family*, 36(1) 57-63. <https://doi.org/10.2307/350994>

- Kluwer, E. S., & Johnson, M. D. (2007). Conflict frequency and relationship quality across the transition to parenthood. *Journal of Marriage and Family*, 69(5), 1089–1106. <https://doi.org/10.1111/j.1741-3737.2007.00434.x>
- Li, F., Ren, L., Wang, X., Jin, Y., Ma, Z., Yang, Q., & Wang, D. (2021). A network perspective on marital satisfaction among older couples. Under Review. <https://doi.org/10.21203/rs.3.rs-674200/v1>
- Locke, H. J., & Wallace, K. M. (1959). Short marital adjustment and prediction tests: Their reliability and validity. *Marriage and Family Living*, 21(3), 251–255. <https://doi.org/10.2307/348022>
- Markus, K.A., & Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning* (1st ed.). Routledge. <https://doi.org/10.4324/9780203501207>
- McWilliams, L. A., & Fried, E. I. (2019). Reconceptualizing adult attachment relationships: A network perspective. *Personal Relationships*, 26(1), 21-41. <https://doi.org/10.1111/pere.12263>
- Nickull, S., Källström, M., & Jern, P. (2022). An exploratory network analysis of sexual and relationship satisfaction comparing partnered cisgendered men and women. *The Journal of Sexual Medicine*, 19(5), 711–718. <https://doi.org/10.1016/j.jsxm.2022.02.006>
- Norton, R. (1983). Measuring marital quality: A critical look at the dependent variable. *Journal of Marriage and the Family*, 45(1), 141. <https://doi.org/10.2307/351302>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 1-10.

- Pierce, G. R., Sarason, I. G., & Sarason, B. R. (1991). General and relationship-based perceptions of social support: Are two constructs better than one? *Journal of Personality and Social Psychology*, *61*(6), 1028–1039. <https://doi.org/10.1037/0022-3514.61.6.1028>
- Ponti, L., Guarnieri, S., Smorti, A., & Tani, F. (2010). A measure for the study of friendship and romantic-relationship quality from adolescence to early-adulthood. *The Open Psychology Journal*, *3*(1). <https://doi.org/10.2174/1874350101003010076>
- Proulx, C., Helms, H., & Buehler, C. (2007). Marital quality and personal well-being: A meta-analysis. *Journal of Marriage and The Family*, *69*, 576–593. <https://doi.org/10.1111/j.1741-3737.2007.00393.x>
- R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reblin, M., Vaughn, A. A., Birmingham, W. C., Smith, T. W., Uchino, B. N., & Spahr, C. M. (2020). Complex assessment of relationship quality within dyads. *Journal of Community Psychology*, *48*(7), 2221–2237. <https://doi.org/10.1002/jcop.22392>
- Roach, A., Frazier, L., & Bowden, S. (1981). The Marital Satisfaction Scale: Development of a measure for intervention research. *Journal of Marriage and Family*, *43*(3), 537-546.
- Robinaugh, D. J., Millner, A. J., & McNally, R. J. (2016). Identifying highly influential nodes in the complicated grief network. *Journal of Abnormal Psychology*, *125*(6), 747–757. <https://doi.org/10.1037/abn0000181>
- Robles, T. F., Slatcher, R. B., Trombello, J. M., & McGinn, M. M. (2014). Marital quality and health: A meta-analytic review. *Psychological Bulletin*, *140*(1). <https://doi.org/10.1037/a0031859>

RStudio Team. (2020). RStudio: Integrated Development Environment for R. Boston, MA.

Retrieved from <http://www.rstudio.com/>

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B., & Trivedi, M. H. (1996). The inventory of depressive symptomatology (IDS): psychometric properties. *Psychological Medicine*, 26(3), 477-486. DOI: 10.1017/s0033291700035558

Rust, J., Bennun, I., Crowe, M., & Golombok, S. (1986). The Golombok Rust inventory of marital state (GRIMS). *Sexual and Marital Therapy*, 1(1), 55-60.

<https://doi.org/10.1080/02674658608407680>

Sabatelli, R. (1984). The Marital Comparison Level Index: A measure for assessing outcomes relative to expectations. *Journal of Marriage and Family*, 46(3), 651-662. <https://doi.org/10.2307/352606>

[10.2307/352606](https://doi.org/10.2307/352606)

Sabatelli, R. (1988). Measurement issues in marital research: A review and critique of contemporary survey instruments. *Journal of Marriage and Family*, 50(4), 891-915.

<https://doi.org/10.2307/352102>

Santor, D. A., Gregus, M., & Welch, A. (2006). Eight decades of measurement in depression measurement: *Interdisciplinary Research and Perspectives*, 4(3), 135-155. [https://doi.org/10.1207/s15366359mea0403\\_1](https://doi.org/10.1207/s15366359mea0403_1)

[doi.org/10.1207/s15366359mea0403\\_1](https://doi.org/10.1207/s15366359mea0403_1)

Schimmack, U. (2021). The validation crisis in psychology. *Meta-Psychology*, 5. <https://doi.org/10.15626/MP.2019.1645>

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, 8(4), 350-353. <https://doi.org/10.1037/1040-3590.8.4.350>

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological

phenomena. *New Ideas in Psychology*, 31(1), 43–53.

<https://doi.org/10.1016/j.newideapsych.2011.02.007>

Schumm, W. R., Paff-Bergen, L. A., Hatch, R. C., Obiorah, F. C., Copeland, J. M., Meens, L. D., & Bugaighis, M. A. (1986). Concurrent and discriminant validity of the Kansas Marital Satisfaction Scale. *Journal of Marriage and the Family*, 381-387.

<https://doi.org/10.2307/352405>

Shah, A. (1995). Clinical validity of Marital Quality Scale. *Nimhans Journal*, 13, 23–21.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha.

*Psychometrika*, 74, 107-120. DOI: 10.1007/s11336-008-9101-0

Silan, M. (2022). *Measuring social constructions? On the difference between indexing and measurement*. Unpublished manuscript. Psychology Department, Universite Lumiere

Lyon 2. Retrieved from: <https://osf.io/5dwvz/>

Silan, M. (2023). *A multi-country investigation into the conceptualization and measurement of relationship quality*. Unpublished manuscript. Psychology Department, Université

Lumiere Lyon 2. Retrieved from: <https://osf.io/n7gu2/>

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, 12(6), 1123-

1128. <https://doi.org/10.1177/1745691617708630>

Snijders, T. A. (2005). Models for longitudinal network data. In P. J. Carrington, J. Scott, & S.

Wasserman (Eds.). *Models and methods in social network analysis* (pp. 215 – 247).

Cambridge University Press.

- Sparacio, A., Ropovik, I., Jiga-Boy, G., & IJzerman, H. (2023a). Stress regulation via being in nature and social support in adults: a meta-analysis. Available via Psyarxiv: <https://psyarxiv.com/a4zmj/>.
- Sparacio, A., Ropovik, I., Jiga-Boy, G., Forscher, P. S., Paris, B., & IJzerman, H. (2023b). Stress regulation via self-administered mindfulness and biofeedback interventions in adults: A pre-registered meta-analysis. Available via PsyArxiv: <https://psyarxiv.com/zpw28/>.
- Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and the Family*, 38(1), 15. <https://doi.org/10.2307/350547>
- Spanier, G. B., & Cole, C. L. (1976). Toward clarification and investigation of marital adjustment. *International Journal of Sociology of the Family*, 6(1), 121–146. <https://doi.org/10.2307/350547>
- Stanley, S. M., Rhoades, G. K., Amato, P. R., Markman, H. J., & Johnson, C. A. (2010). The timing of cohabitation and engagement: Impact on first and second marriages. *Journal of Marriage and Family*, 72(4), 906–918 <https://doi.org/10.1111/j.1741-3737.2010.00738.x>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53. DOI: 10.5116/ijme.4dfb.8dfd
- Thorndike, E. L. (1904). *Theory of mental and social measurements*. The Science Press. <https://doi.org/10.1037/13283-000>
- Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-augmented meta-analyses: Toward cumulative data assessment. *Perspectives on Psychological Science*, 9(6), 661-665. <https://doi.org/10.1177/1745691614552498>

- Visontay, R., Sunderland, M., Grisham, J., & Slade, T. (2019). Content overlap between youth OCD scales: Heterogeneity among symptoms probed and implications. *Journal of Obsessive-Compulsive and Related Disorders*, 21, 6–12.  
<https://doi.org/10.1016/j.jocrd.2018.10.005>
- Wahyuningsih, H., Kusumaningrum, F. A., & Novitasari, R. (2020). Parental marital quality and adolescent psychological well-being: A meta-analysis. *Cogent Psychology*, 7(1), 1819005. <https://doi.org/10.1080/23311908.2020.1819005>
- Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, 17(2), 267–295. <https://doi.org/10.1037/emo0000226>
- Wetzel, E., & Roberts, B. W. (2020). Commentary on Hussey and Hughes (2020): Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*, 3(4), 505-508. <https://doi.org/10.1177/2515245920957618>
- Wilhelm, K., & Parker, G. (1988). Reliability of the parental bonding instrument and intimate bond measure scales. *Australian & New Zealand Journal of Psychiatry*, 24(2), 199-202.  
<https://doi.org/10.1017/S0033291700002051>
- Zhang, H., Xu, X., & Tsang, S. K. (2013). Conceptualizing and validating marital quality in Beijing: A pilot study. *Social Indicators Research*, 113(1), 197-212. <https://doi.org/10.1007/s11205-012-0089-6>

**Study 2's Appendices**

Appendix A: Maps of participants' living locations

Appendix B: Full List of Romantic Relationship Items

Appendix C: Stability Analysis of the 33-Node Network

Appendix D: Dimensions and Items Stability of the 33-node Network

Appendix E: Stability Analysis of the 24-Node Network

Appendix F: Dimensions and Items Stability of the 24-node Network



### Figure Legends

**Figure 1.** *Flowchart of the Scale Selection (Search 1)*

**Figure 2.** *Origin of the 26 self-reported scales (Scales per country).*

**Figure 3.** *Co-occurrence of 25 thematic categories across 26 self-reported scales. Figure Legend. Co-occurrence of 25 thematic categories across 25 romantic relationship quality scales. Squares for a category indicate that a scale only measures the category indirectly, circles indicate that a scale only measures the category directly, and triangles indicate that a scale measures the category both indirectly and directly. Shapes (circles, squares, and triangles) represent scales (from outer to inner shapes, with the number of categories assessed): CMQI (22), RRF (21), MQS (21), QDR (19), MSS (19), CMQS (19), MCLI (18), RMAT (16), MIRS (16), GRIMS (16), DAS (16), MMQ (14), MAT (14), QRI (12), IRQ (12), IBM (12), GLRSS (12), RQS (11), PBSC (11), RDAS (10), CSI (10), PRQC (6), RAS (5), QMI (4), PANQIMS (2), KMSS (1).*

**Figure 4.** *Correlation heatmap of overlap of item content of 26 romantic-relationship quality scales.*

### Figure 5

*(A) Regularized Network of 33 Romantic Relationship Quality Items, (B) Centrality Indices and (C) Items Stability*

*Figure legend 5A. Gaussian graphical model of the 33 items. Colors of the nodes distinguish the four detected communities.*

*Figure Legend 5B. Expected influence of the 33 items. Expected influence of a given node is the standardized sum of the weights of all direct connections between a specific item and all other items. Estimates are transformed z-score. Higher scores reflect a higher degree of importance of connectiveness within a network.*

*Figure Legend 5C. Number of times each item replicated within the empirical EGA defined dimension.*

**Figure 6**

*Regularized Network of 24 Romantic Relationship Quality Items, (B) Items Stability, (C) Node Expected Influence, and (D) Bridge Expected Influence.*

*Figure Legend 6A. Gaussian graphical model of the 24 items. Colours of the nodes denote four detected communities.*

*Figure Legend 6B. Number of times each item replicated within the empirical EGA defined dimension.*

*Figure Legend 6C. Expected influence of the 24 items. Expected influence of a given node is the standardized sum of the weights of all direct connections.*

**Table 1.** *Summary of Scales' Characteristics*

Scale	Authors	Date	Country	Items
CMQI	Cheng et al.	2010	China	90
CMQS	Zhang et al.	2012	China	39
CSI	Funk & Rogge	2007	USA	32
DAS	Spanier	1976	USA	32
GLRSS	Belous & Wampler	2016	USA	24
GRIMS	Rust et al.	1986	UK	28
IBM	Wilhelm & Parker	1988	Australia	24
IRQ	Braiker & Kelley	1979	USA	25
KMSS	Schumm et al.	1986	USA	3
MAT	Locke & Wallace	1959	USA	15
MCLI	Sabatelli	1984	USA	36
MIRS	Garthoeffner et al.	1993	USA	46
MMQ	Arrindell et al.	1983	Netherlands	20
MQS	Shah	1995	India	50
MSS	Blum & Mehrabien	1999	USA	35
PANQIMS	Fincham & Linfield	1997	UK	6
PBSC	Ducat & Zimmer-Gembeck	2010	Australia	30
PRQC	Fletcher et al.	2000	New Zealand	18
QDR	Ahlborg et al.	2009	Sweden	36
QMI	Norton	1983	USA	6
QRI	Pierce et al.	1991	USA	25
RAS	Hendrick	1988	USA	7
RDAS	Busby et al.	1995	USA	14
RMAT	Kimmel & Van Der Veen	1974	USA	23
RQS	Ponti et al.	2010	Italy	22
RRF	Davis	2001	USA	68

*Note.*

CMQI: Chinese Marital Quality Inventory;  
 CMQS: Chinese Marital Quality Scale;  
 CSI: Couples Satisfaction Index;  
 DAS: Dyadic Adjustment Scale;  
 GLRSS: Gay and Lesbian Relationship Satisfaction Scale;  
 GRIMS: Golombok-Rust Inventory of Marital State;  
 IBM: Intimate Bond Measure;  
 IRQ: Intimate Relations Questionnaire;  
 KMSS: Kansas Marital Satisfaction Scale;  
 MAT: Marital Adjustment Test;  
 MCLI: Marital Comparison Level Index;  
 MIRS: Modified Interpersonal Relationship Scale;  
 MMQ: Maudsley Marital Questionnaire;  
 MQS: Marital Quality Scale;

MSS: Marital Satisfaction Scale;  
 PANQIMS: Positive and Negative Quality in Marriage Scale;  
 PBSC: Partner Behaviors as Social Context scale;  
 PRQC: Perceived Relationship Quality Components inventory;  
 QDR: Quality of Dyadic Relationship;  
 QMI: Quality of Marriage Index;  
 QRI: Quality of Relationship Inventory  
 RAS: Relationship Assessment Scale;  
 RDAS: Revised Dyadic Adjustment Scale;  
 RMAT: Revised Marital Adjustment Test;  
 RQS: Romance Qualities Scale;  
 RRF: Relationship Rating Form.

**Table 2.** *Occurrence of categories and their number of items*

Category	Appears in N scales	Composed of N items
Affection/Love	22	68
Emotional Closeness/Intimacy	20	61
Global Happiness/Satisfaction	20	57
Tension/Conflict	19	60
Compatibility in Attitudes/Preferences	19	60
Oral Communication	18	47
Commitment/Involvement	17	50
Sexuality	15	41
Time spent together	15	37
Expectations/Needs met	15	34
Trust	14	68
Emotional Support	14	63
Conflict resolution	14	32
Stability	14	26
Consensus in decision making	13	54
Social Networks	12	44
Shared Life Philosophies	12	34
Power/Control	11	54
Mutual Dependence	11	28
Instrumental Support	11	25
Exclusiveness/Monogamy	10	21
Physical Affection	9	13
Family Life/Parenting	7	26
Agreement about proper conduct	7	8
Forgiveness	1	2
Miscellaneous items	2	2

**Table 3.** *Categories per scale and general vs. specific items*

Scale	Number of scale items:	Captured Category	% of all 25 disparate categories captured	Idiosyncratic Category	% of specific items	% of general items
CMQI	90	22	88	0	75.56	24.44
CMQS	39 (38)*	19	76	0	36.84	63.16
CSI	32	10	40	0	53.12	46.88
DAS	32	16	64	0	43.75	56.25
GLRSS	24	12	48	0	79.17	20.83
GRIMS	28	16	64	0	71.43	28.57
IBM	24	12	48	0	41.67	58.33
IRQ	25	12	48	0	48.00	52.00
KMSS	3	1	4	0	0	100.00
MAT	15	14	56	0	53.33	46.67
MCLI	36	18	72	0	50.00	50.00
MIRS	46	16	64	0	76.09	23.91
MMQ	20 (19)*	14	56	0	42.11	57.89
MQS	50	21	84	0	66.00	34.00
MSS	35	19	76	0	42.86	57.14
PANQIMS	6	2	8	0	100.00	0
PBSC	30	11	44	0	50.00	50.00
PRQC	18	6	24	0	11.11	88.89
QDR	36	19	76	0	52.78	47.22
QMI	6	4	16	0	33.33	66.67
QRI	25	12	48	0	68.00	32.00
RAS	7	5	20	0	28.57	71.43
RDAS	14	10	40	0	35.71	64.29
RMAT	23	16	64	0	34.78	65.22
RQS	22	11	44	0	59.09	40.91
RRF	68	21	84	1	58.82	41.18
mean		13.04	52.15		55.85	44.15

\*The values in parentheses in the original scale lengths are the adjusted lengths of the scales. The CMQS and MMQ each have a "miscellaneous" item; Explanation of *Captured category*: number of categories out of 25 disparate categories that the scale captures; Explanation of *Idiosyncratic category*: number of categories that appear in no other scale; Explanation of *% of specific items*: percentage of items that indirectly measure the categories; Explanation of *% of general items*: percentage of items that directly measure the categories. All results and data are available at <https://osf.io/c6zww>.

**Table 4.** Validity evidence reported in 43 scale development-validation articles

Scales	Factor analyses		Internal consistency		Test-retest reliability	Measurement invariance
	EFA/PCA	CFA	Alpha	Omega		

---

CMQI	No	Yes	Yes	No	Yes	No
	No	No	Yes	No	No	No
CMQS	Yes	Yes	Yes	No	No	No
CSI	Yes	No	Yes	No	No	No
	No	No	Yes	No	No	No
DAS	No	No	Yes	No	No	No
	Yes	No	Yes	No	No	No
GLRSS	Yes	Yes	Yes	No	No	No
	Yes	Yes	Yes	No	No	No
GRIMS	Yes	No	Yes	No	No	No
	No	No	Yes	No	Yes	No
IBM	Yes	Yes	Yes	No	Yes	No
	Yes	No	Yes	No	No	No
IRQ	No	No	Yes	No	No	No
	Yes	No	No	No	No	No
KMSS	No	Yes	Yes	No	No	Yes
	No	No	Yes	No	Yes	No
MAT	Yes	No	Yes	No	Yes	No
	No	No	No	No	No	No
MCLI	Yes	No	Yes	No	No	No
	Yes	No	Yes	No	Yes	No
MIRS	Yes	No	Yes	No	No	No
	No	No	Yes	No	No	No
MMQ	Yes	Yes	Yes	No	Yes	Yes
	Yes	No	Yes	No	No	No
MQS	Yes	No	Yes	No	Yes	No
	No	No	Yes	No	No	No
MSS	No	No	Yes	No	No	No
	Yes	No	Yes	No	Yes	No
PANQIMS	No	Yes	Yes	No	No	No
	No	Yes	Yes	No	No	No

PBSC	Yes	No	Yes	No	No	No
	No	No	Yes	No	No	No
QDR	Yes	No	Yes	No	No	No
	No	No	Yes	No	No	No
QMI	No	Yes	Yes	No	No	No
	Yes	No	No	No	No	No
QRI	Yes	No	Yes	No	No	No
RAS	No	Yes	Yes	No	Yes	No
	Yes	No	Yes	No	No	No
RMAT	Yes	No	No	No	Yes	No
RQS	No	Yes	Yes	No	No	Yes
	No	No	Yes	No	No	No
Evidence reported	23	12	39	0	11	3
% of articles that had reported evidence	53.49%	27.91%	90.70%	0%	25.58%	6.98%

**Table 5***Characteristics of Sample*

	<i>N(%)</i>	<i>M(SD)</i>
Age		32.4(12.9)
Gender		
Female	486(82.8)	
Male	90(15.3)	
Non-binary/Third gender	9(1.5)	
Unidentified	2 (0.3)	
Relationship status		
Seriously dating	412(70.2)	
Engaged	34(5.8)	
Married	141(24)	

	<i>N</i> (%)	<i>M</i> ( <i>SD</i> )
Relationship duration		
6 months to less than 1 year	85(14.5)	
1 year to less than 3 years	160(27.3)	
3 years to less than 5 years	109(18.6)	
5 years to less than 10 years	89(15.2)	
10 years or more	144(24.5)	
Sexual orientation		
Heterosexual	483(82.3)	
Homosexual	22(3.7)	
Bisexual	69(11.8)	
Asexual	5(0.9)	
Other	8(1.4)	
Monogamy in current relationship <sup>a</sup>		1.22(0.68)
Monogamy in ideal relationship <sup>a</sup>		1.42(0.87)
Have children <sup>b</sup>	212(36.1)	
Have children with current partner <sup>b</sup>	144(24.5)	

<sup>a</sup> Monogamy in current and ideal relationship was scored on 5-point scales (ranging from 1 being “Entirely monogamous” to 5 being “Entirely polyamorous”)

<sup>b</sup> Reflects the number and percentage of participants answering “yes” to these questions.

**Table 6**

*Items Description and Abbreviation*

Node Abbreviation	Item
Forgive	It is easy for me to forgive my partner
NotLoved	I don't get the love and affection I want from my partner
Love	I love my partner
Personalities	Our personalities are rather matched
PerfectPart	For me, my partner is the perfect romantic partner
TalkFight	When there is a difference of opinion, we try to talk it out rather than fight
ImpDecisions	My partner and I agree to make important decisions
Conv	I sometimes worry about his/her behaviours on social occasions
StrgConn	I have a strong connection with my partner



Enc	My partner encourages me to do the things I think are important
OnlyPart	There are things I do only with my partner
Needs	This relationship with my partner satisfies my need
Home	I'm happy with the way my partner takes care of things at home
Needed	My partner makes me feel needed
Depend	I often depend on my partner
ComSat	I am very happy with the way we talk and communicate
Hug	I often want to hug and cuddle my partner
Control	My partner often tries to control or influence my life
OwnCommit	I am committed to our relationship
PartCommit	My partner is committed to our relationship
GoingWell	Things are often going well between us
Happy	My relationship with my partner makes me happy
Stable	My relationship with my partner is very stable
Separate	We often discuss or consider divorce, separation, or terminating our relationship
Sex	I am happy with our sexual relationship
DiffViews	We have large differences in our views towards life
Inlaws	My partner and I agree on our dealings with our in-laws
Friends	My partner and I agree regarding friends
Quarrel	We often quarrel
Angry	I often feel angry or resentful toward my partner
Outside	My partner and I often engage in outside interests together
Fun	My partner and I often have fun together
Trust	I trust my partner

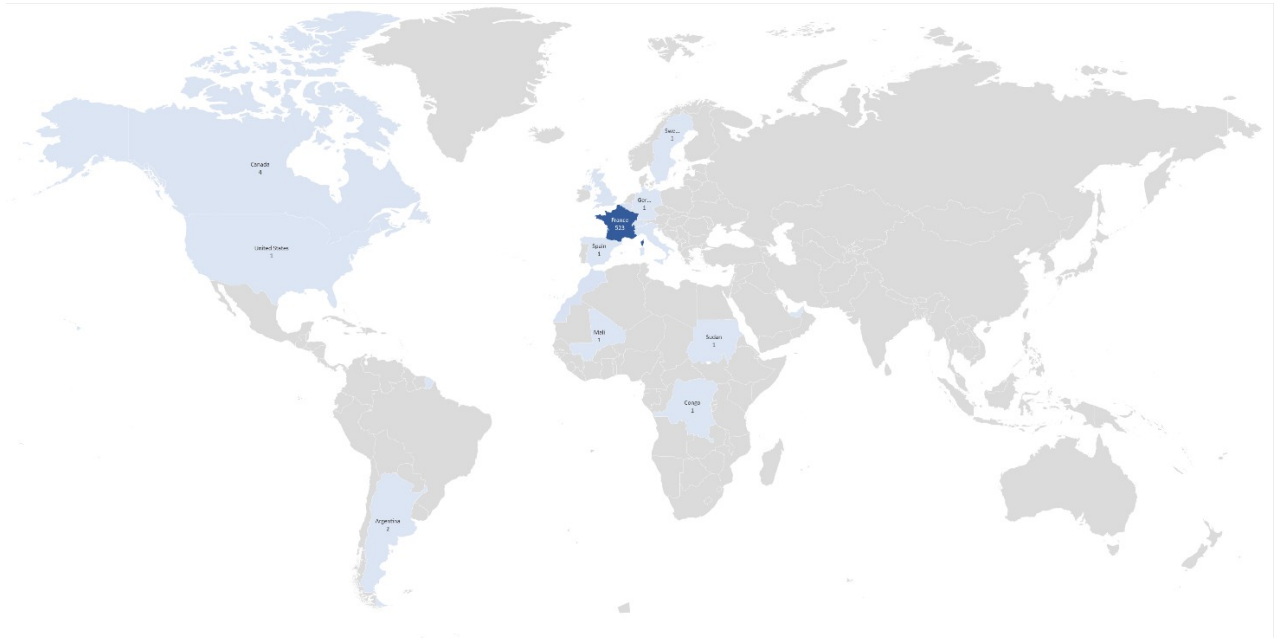
---

## Appendix A

### Maps of participants' living locations

#### Figure A1

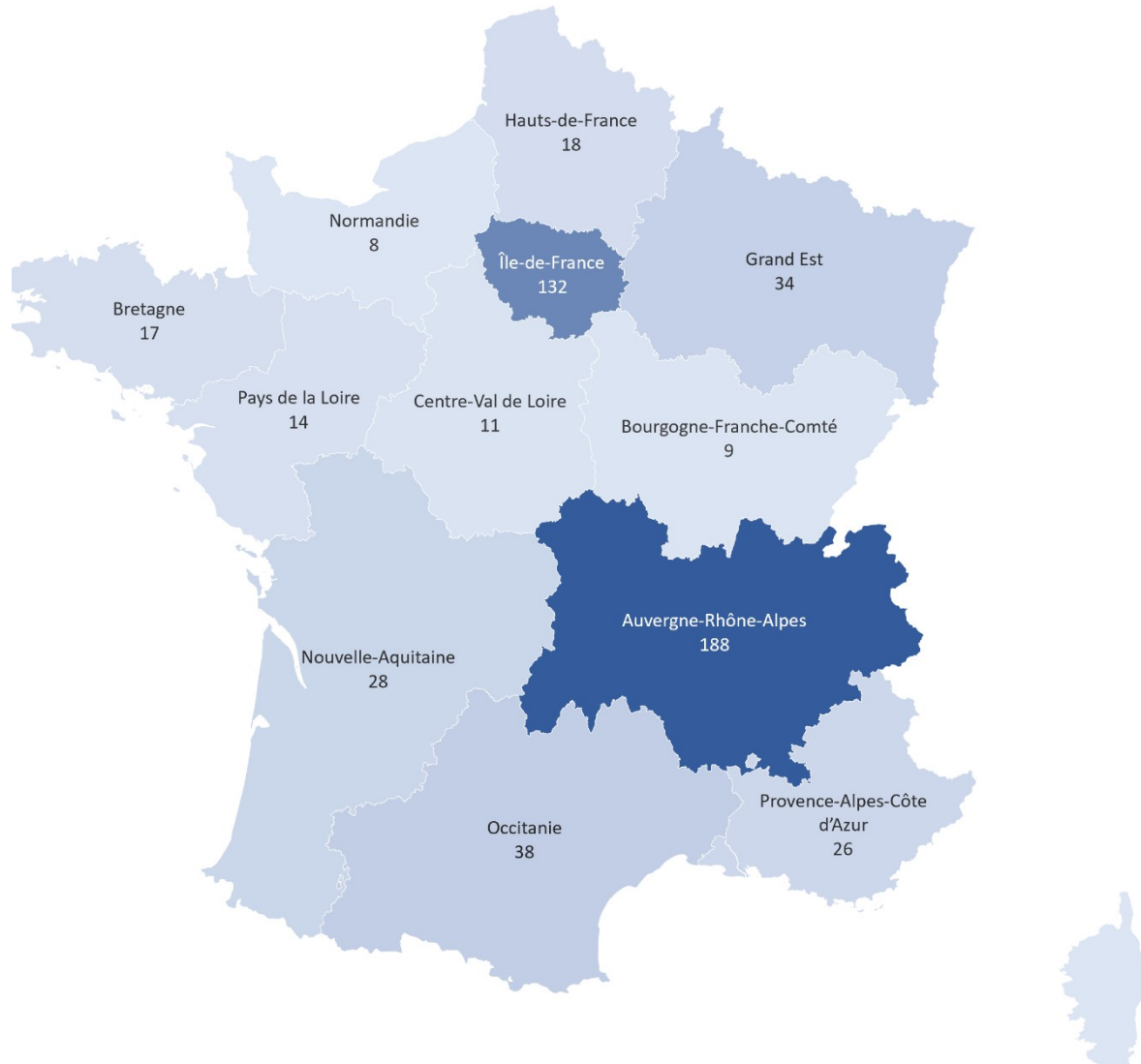
*Participants' places of living by country.*



*Note.* The darker the shade of blue, the more participants the country represents. The number under the country name represents the number of participants who indicated that they live in that country. Of a total of 587 participants, a considerable proportion responded that they lived in mainland France (N = 523); 7 participants did not respond. The other countries, as well as French overseas departments and regions, are the following: Belgium (14); Morocco (11); United Kingdom (5); Canada (4); Switzerland (4); Réunion (3); Argentina (2); French Guiana (2); Congo (1); Germany (1); Guadeloupe (1); Italy (1); Mali (1); Mayotte (1); Spain (1); Sudan (1); Sweden (1); United Arab Emirates (1); United States (1).

### **Figure A2.**

*Participants' places of living by region on mainland France.*



*Note.* The darker the shade of blue, the more participants the region represents. The number under the province name represents the number of participants who indicated that they live in that region. A total of 523 participants responded that they live in one of these regions.

## Appendix B

### Full List of Romantic Relationship Items

**Table B1.**

*The initial 50 items selected measuring romantic-relationship quality*

Category associated	Description	French translation
Acceptance/Forgiveness	<b>It is easy for me to forgive my partner</b>	Il m'est facile de pardonner mon partenaire (RQ_1)
	I'm willing to ignore my partner's small sins because of the way I feel about him/her	Je suis prêt(e) à ignorer les petits défauts de mon partenaire en raison de ce que je ressens pour lui/elle (RQ_2)
Affection/Love	<b>I don't get the love and affection I want from my partner</b>	Je n'obtiens pas de mon partenaire l'amour et l'affection que je désire (RQ_3)
	<b>I love my partner</b>	J'aime mon partenaire (RQ_4)
Compatibility	<b>Our personalities are rather matched</b>	Nos personnalités sont plutôt compatibles (RQ_5)
	<b>For me, my partner is the perfect romantic partner</b>	Pour moi, mon partenaire est le partenaire romantique idéal (RQ_6)
Conflict resolution	<b>When there is a difference of opinion, we try to talk it out rather than fight</b>	Lorsqu'il y a une différence d'opinion, nous essayons d'en parler plutôt que de nous disputer (RQ_7)
	When we have arguments, we are able to reach a compromise	Lorsque nous nous disputons, nous parvenons à trouver un compromis (RQ_8)
Consensus in decision making	<b>My partner and I agree to make important decisions</b>	Mon partenaire et moi sommes d'accord pour prendre des décisions importantes (RQ_9)
	My partner and I agree on how we handle our finances	Mon partenaire et moi sommes d'accord sur la façon dont nous gérons nos finances (RQ_10)
Conventionality	My partner and I agree on conventionality (correct or proper behavior during social activities )	Mon partenaire et moi sommes d'accord sur les conventions (comportement correct ou approprié lors d'activités sociales) (RQ_11)
	<b>I sometimes worry about his/her behaviours on social occasions</b>	Je m'inquiète parfois de son comportement en société (RQ_12)
Emotional Closeness/Intimacy	<b>I have a strong connection with my partner</b>	J'ai un lien fort avec mon partenaire (RQ_13)
	I often confide in my partner	Je me confie souvent à mon partenaire (RQ_14)
Emotional Support	My partner supports my interests	Mon partenaire soutient mes intérêts (RQ_15)
	<b>My partner encourages me to do the things I think are important</b>	Mon partenaire m'encourage à faire les choses que je juge importantes (RQ_16)
Exclusiveness/Monogamy	In the past year, me and my partner has often been jealous	Au cours de l'année écoulée, mon partenaire a souvent été jaloux(se) (RQ_17)
	<b>There are things I do only with my partner</b>	Il y a des choses que je ne fais qu'avec mon

ASSESSMENT OF RELATIONSHIP QUALITY 76

Category associated	Description	French translation
		partenaire (RQ_18)
Expectations/Needs met	My partner disappoints me <b>This relationship with my partner satisfies my needs</b>	Mon partenaire me déçoit (RQ_19) La relation avec mon partenaire satisfait mes besoins (RQ_20)
Instrumental Support	<b>I'm happy with the way my partner takes care of things at home</b>  I can count on my partner to lend me a substantial amount of money	Je suis satisfait de la façon dont mon partenaire s'occupe des choses à la maison (RQ_21)  Je peux compter sur mon partenaire pour me prêter une somme d'argent importante (RQ_22)
Mutual Dependence	<b>My partner makes me feel needed</b>  <b>I often depend on my partner</b>	Mon partenaire me fait sentir qu'on a besoin de moi (RQ_23)  Je dépends souvent de mon partenaire (RQ_24)
Oral Communication	<b>I am very happy with the way we talk and communicate</b> We often have a stimulating exchange of ideas	Je suis très satisfait(e) de la façon dont nous parlons et communiquons (RQ_25) Nous avons souvent des échanges d'idées stimulants. (RQ_26)
Parenting	I am not happy with my position in the family  I am happy with the responsibilities we share in raising our child/ren	Je ne suis pas satisfait de ma position dans la famille (RQ_27)  Je suis satisfait(e) des responsabilités que nous partageons dans l'éducation de nos enfants (RQ_28)
Physical Affection	<b>I often want to hug and cuddle my partner</b>  I enjoy being touched by my partner and touching him/her	J'ai souvent envie d'étreindre et de câliner mon partenaire (RQ_29)  J'aime être touché(e) par mon/ma partenaire et le/la toucher (RQ_30)
Power/Control	My partner tends to control everything I do  <b>My partner often tries to control or influence my life</b>	Mon partenaire a tendance à contrôler tout ce que je fais (RQ_31)  Mon partenaire essaie souvent de contrôler ou d'influencer ma vie (RQ_32)
Relationship Commitment	<b>I am committed to our relationship</b>  <b>My partner is committed to our relationship</b>	Je suis engagé(e) dans notre relation (RQ_33)  Mon partenaire est engagé dans notre relation (RQ_34)
Relationship Satisfaction	<b>Things are often going well between us</b>  <b>My relationship with my partner makes me happy</b>	Les choses se passent souvent bien entre nous (RQ_35)  Ma relation avec mon partenaire me rend heureux(se) (RQ_36)
Relationship Stability	<b>My relationship with my partner is very stable</b>  <b>We often discuss or consider divorce, separation, or terminating our relationship</b>	Ma relation avec mon/ma partenaire est très stable (RQ_37)  Nous discutons ou envisageons souvent le divorce, la séparation ou la fin de notre relation (RQ_38)
Sexuality (Desire)	I often feel sexual desire for my partner  <b>I am happy with our sexual relationship</b>	Je ressens souvent du désir sexuel pour mon partenaire (RQ_39)  Je suis heureux(se) de notre relation sexuelle (RQ_40)
Shared Life Philosophies	My partner and I share the same values and goals in life	Mon partenaire et moi partageons les mêmes valeurs et objectifs dans la vie

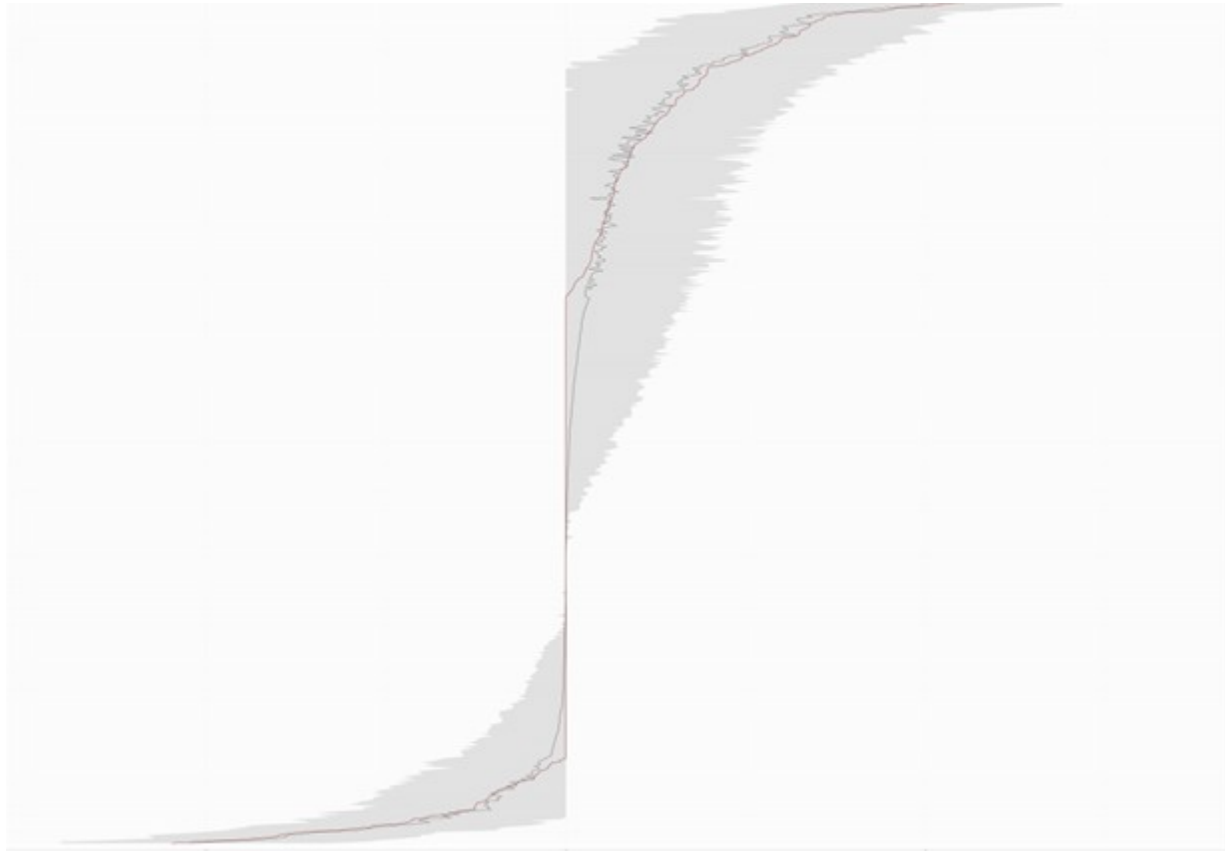
## ASSESSMENT OF RELATIONSHIP QUALITY 77

Category associated	Description	French translation
		(RQ_41)
	<b>We have large differences in our views towards life</b>	Nous avons de grandes différences d'opinion sur la vie (RQ_42)
Social Networks	<b>My partner and I agree on our dealings with our in-laws</b>	Mon partenaire et moi sommes d'accord sur nos relations avec la belle-famille (RQ_43)
	<b>My partner and I agree regarding friends</b>	Mon partenaire et moi sommes d'accord en ce qui concerne les amis (RQ_44)
Tension/Conflict	<b>We often quarrel</b>	Nous nous disputons souvent (RQ_45)
	<b>I often feel angry or resentful toward my partner</b>	Je ressens souvent de la colère ou du ressentiment envers mon partenaire (RQ_46)
Time spent together	<b>My partner and I often engage in outside interests together</b>	Mon partenaire et moi partageons souvent des activités extérieures ensemble (RQ_47)
	<b>My partner and I often have fun together</b>	Mon partenaire et moi nous amusons souvent ensemble (RQ_48)
Trust in the partner	My partner is not someone who I can always rely on	Mon partenaire n'est pas quelqu'un sur qui je peux toujours compter (RQ_49)
	<b>I trust my partner</b>	J'ai confiance en mon partenaire (RQ_50)

*Note.* The initial 50 items we administered to measure romantic relationship quality. In bold are the items used in the network analysis.

**Appendix C****Stability Analysis of the 33-Node Network****Figure C1.**

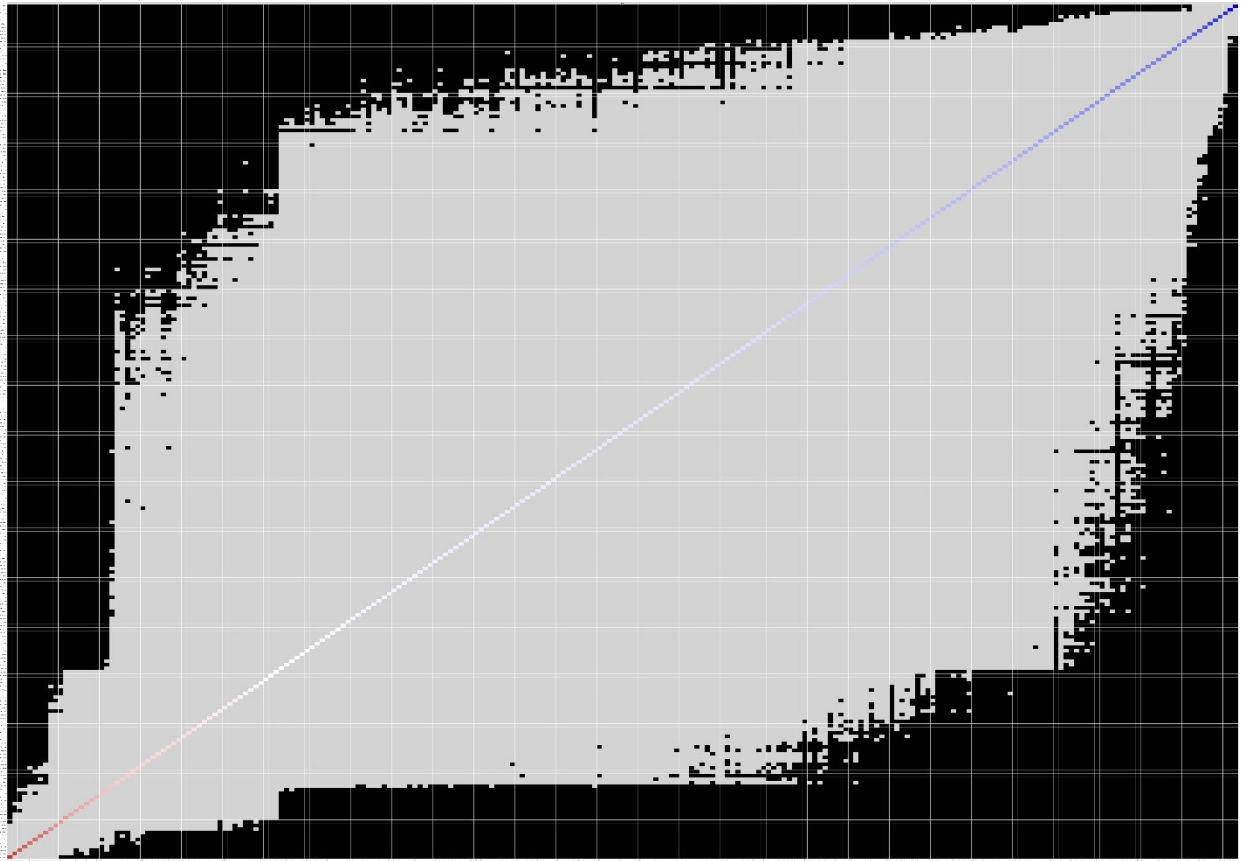
*Bootstrapped confidence intervals of estimated edge-weights*



*Note.* Each edge of the network is represented in one horizontal line and is ordered from the edge with the highest edge weight to the lowest edge weight. The grey area represents the bootstrapped confidence intervals and the red line represents the sample values. The smaller the confidence intervals, the more accurate the edge estimates. Overlapping confidence intervals signify that edge weights are unlikely to differ from one another significantly and that interpreting the order of most edges in the network should be done carefully.

**Figure C2.**

*Bootstrapped difference test of estimated edge weights*

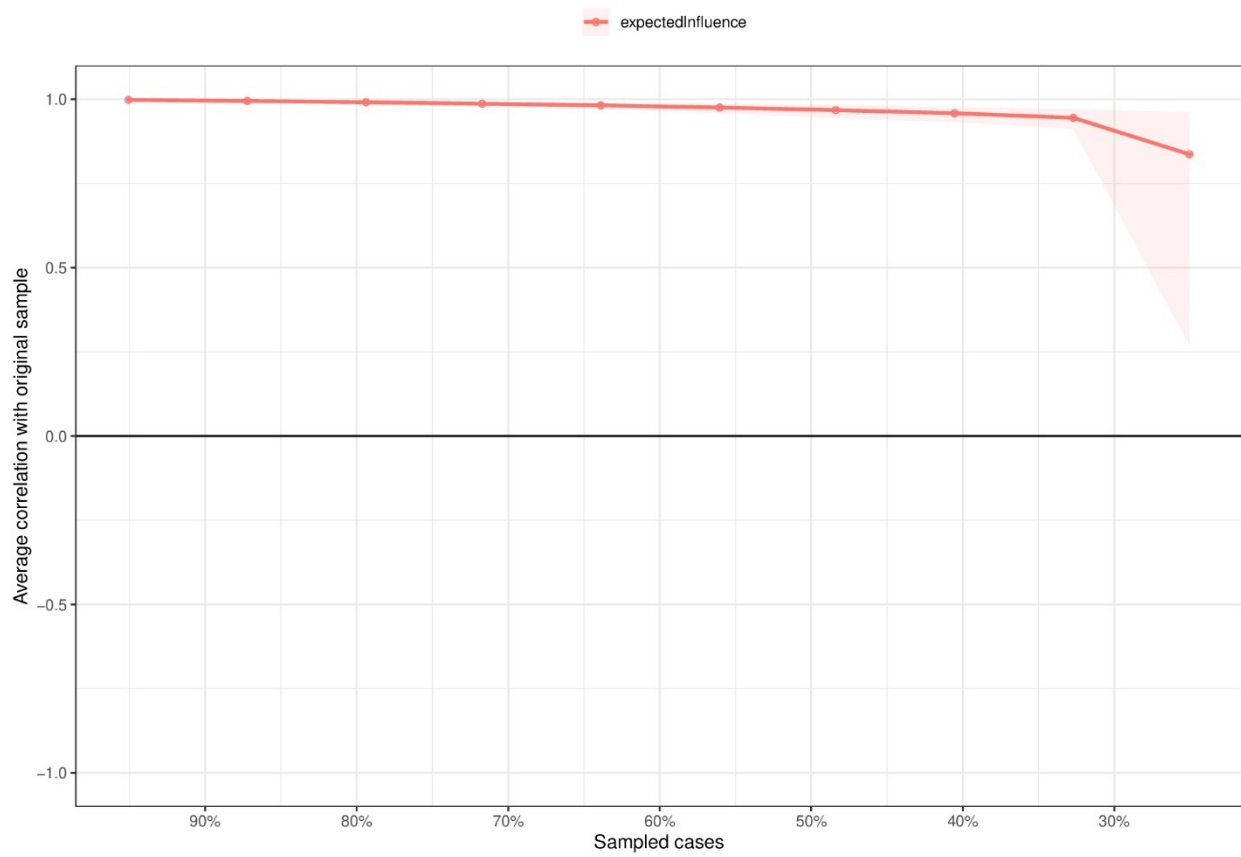


*Note.* Both axes represent all non-zero edges in the network from the lowest to the highest value. Statistical differences between two edge weights are indicated with a black box; non-statistical differences are indicated with a grey box. These results suggest that the strongest (negative and positive) edge-weights comparisons were significantly different from each other ( $p < .05$ ).



**Figure C3.**

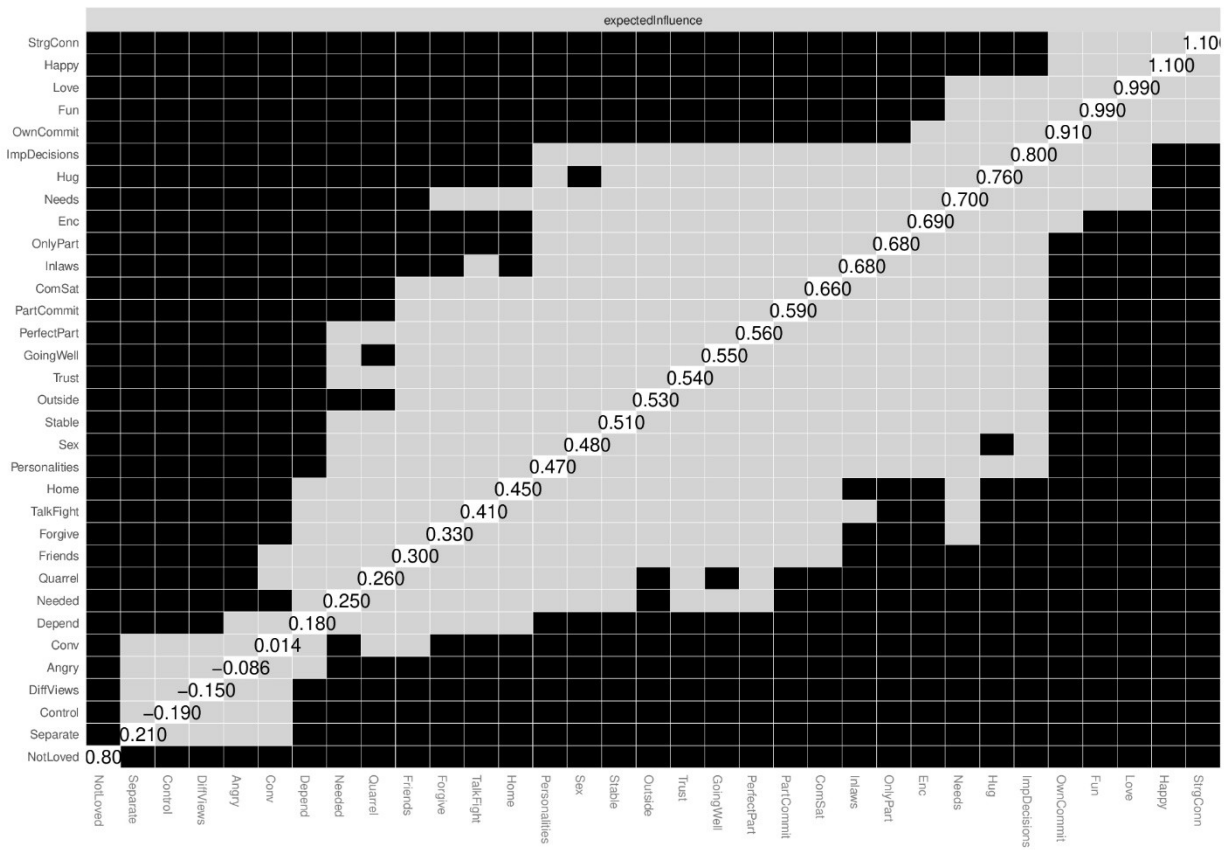
*Stability of estimated expected influence*



*Note.* Subset bootstrapping of the 33 items showing the correlation between the centrality values in the original network and networks estimated with fewer subjects (1,500 bootstrapped samples). The red line indicates the mean and the red areas represent the range from the 2.5th quantile to the 97.5th quantile. CS-coefficient ((CS(cor = 0.7) = 0.75) indicates the mean percentage of our sample that can be dropped to preserve a correlation of  $r = .7$  between our sample's centrality indices and our case-dropped bootstraps' centrality indices. CS-coefficient should not be below 0.25 and preferably above 0.5.

**Figure C4.**

*Bootstrapped difference test of estimated expected influence*



*Note.* Both axes represent all degree centrality estimates for the 33 items of the questionnaire, from the lowest to the highest value. Statistical differences between two centrality estimates are represented by a black box; non-statistical difference are represented by a grey box. These results suggest that the strongest centrality estimates comparisons are significantly different from a large majority of other nodes ( $p < .05$ ).

**Appendix D****Dimensions and Items Stability of the 33-node Network****Table D1.***Stability statistics for the EGA network derived from 1500 bootstrapped samples*

Summary statistics for bootstrapped EGA					
N of boots	Med.dim	SE.dim	CI.dim	Lower	Upper
1500	4	0.433	0.849	3.151	4.849

*Note.* This table provides descriptive statistics for the number of bootstraps, median number of dimensions found across the bootstraps, standard error (SE.dim), and lower (Lower) and upper (Upper) bound of the 95% confidence interval of the bootstrapped dimensions.

**Table D2.***Dimensions likelihood for the EGA network derived from 1500 bootstrapped samples*

# of dimensions	Likelihood
3	0.033
4	0.803
5	0.161
6	0.003

*Note.* Proportion of times the number of dimensions was identified.

**Table D3.***Structural consistency of dimensions and average item stability*

Dimension	Structural consistency	Average item stability
1	0.225	0.727
2	0.289	0.860

3	0.051	0.809
4	0.941	0.733

*Note.* Structural consistency is the proportion of times that each empirical EGA dimension exactly replicates across the 1,500 bootstrapped samples. We also provided the average item stability in each empirical EGA dimension.

**Table D4.**

*Item likelihood for the EGA network derived from 1,500 bootstrapped samples*

	Dimensions					
	1	2	3	4	5	6
Inlaws	<b>0.914765101</b>	0.009395973	0.02885906	0.00738255	0.038926174	0.000671141
ImpDecisions	<b>0.883658373</b>	0.036314728	0.030262273	0.002017485	0.047074647	0.000672495
Friends	<b>0.872155288</b>	0	0.081659973	0	0.045515395	0.000669344
Trust	<b>0.840696118</b>	0.000669344	0.093708166	0.011378849	0.052878179	0.000669344
DiffViews	<b>0.733779264</b>	0.000668896	0.222073579	0.002675585	0.040802676	0
Enc	<b>0.698049765</b>	0.061197041	0.176193679	0.016139879	0.047747142	0.000672495

ASSESSMENT OF RELATIONSHIP QUALITY 86

Conv	<b>0.679812834</b>	0	0.270721925	0.005347594	0.043449198	0.000668449
Forgive	<b>0.670020121</b>	0.036217304	0.183098592	0.072434608	0.037558685	0.000670691
Home	<b>0.622654155</b>	0.018766756	0.18230563	0.13538874	0.040884718	0
Personalities	<b>0.608026756</b>	0.013377926	0.331772575	0.004682274	0.042140468	0
Hug	0.002	<b>0.992666667</b>	0	0.004	0.001333333	0
Love	0.006048387	<b>0.987231183</b>	0	0	0.006048387	0.000672043
StrgConn	0.006048387	<b>0.987231183</b>	0	0	0.006048387	0.000672043
OwnCommit	0.006048387	<b>0.987231183</b>	0	0	0.006048387	0.000672043
OnlyPart	0.02014775	<b>0.957689725</b>	0.001343183	0.010073875	0.009402283	0.001343183
Sex	0.001333333	<b>0.887333333</b>	0	0.109333333	0.002	0
Fun	0.001333333	<b>0.884</b>	0.001333333	0.102	0.011333333	0
Outside	0.022712091	<b>0.817635271</b>	0.014696059	0.12758851	0.017368069	0
Happy	0.006004003	<b>0.518345564</b>	0.080053369	0.345563709	0.049366244	0.000667111
Needs	0	0.325333333	0.011333333	<b>0.638666667</b>	0.024666667	0
Quarrel	0.039333333	0	<b>0.94</b>	0	0.020666667	0
Angry	0.039359573	0	<b>0.939959973</b>	0	0.020680454	0
GoingWell	0.031333333	0.006	<b>0.878</b>	0.04	0.044666667	0
ComSat	0.08811749	0	<b>0.872496662</b>	0.000667557	0.038718291	0
TalkFight	0.090787717	0	<b>0.869826435</b>	0.000667557	0.038718291	0
Separate	0.209893048	0.000668449	<b>0.577540107</b>	0.137700535	0.072860963	0.001336898
Control	0.428475936	0.012032086	<b>0.521390374</b>	0.001336898	0.036764706	0
Stable	0.253342246	0.002005348	<b>0.520721925</b>	0.145721925	0.076871658	0.001336898
Depend	0.276209677	<b>0.404569892</b>	0.268145161	0.016129032	0.034946237	0
NotLoved	0.002	0.034666667	0	<b>0.96</b>	0.003333333	0
Needed	0.003333333	0.034	0	<b>0.958666667</b>	0.004	0
PartCommit	0.004	0.033333333	0	<b>0.958</b>	0.004666667	0
PerfectPart	0.01	0.039333333	0.002666667	<b>0.946666667</b>	0.001333333	0

*Note.* Item stability. Each value represents the proportion of times each item is replicated in each of the empirical EGA-defined dimensions. We used the input from the EGA network derived from 1500 bootstrapped samples. The highest absolute likelihood for an item is highlighted in bold font.

**Table D5.**

*Average network loading for each item in each dimension over 1,500 bootstrapped samples*

	Dimensions					
	1	2	3	4	5	6
Inlaws	<b>0.139534899</b>	0.049191275	0.01356076	0.039975069	0.05797551	0.07625
ImpDecisions	<b>0.174578346</b>	0.076195696	0.03422585	0.04066759	0.096681633	0.0945
Friends	<b>0.171740964</b>	0.012133869	0.027707317	0.012889273	0.069265306	0.066
Trust	<b>0.189274431</b>	0.018766399	0.050409214	0.048524567	0.094089796	0.007
DiffViews	<b>-0.10226355</b>	-0.003757191	0.016490183	-0.008709343	-0.012514286	-0.02925
Enc	<b>0.111489576</b>	0.070170141	0.045710884	0.051876731	0.05522449	0.0855

## ASSESSMENT OF RELATIONSHIP QUALITY 87

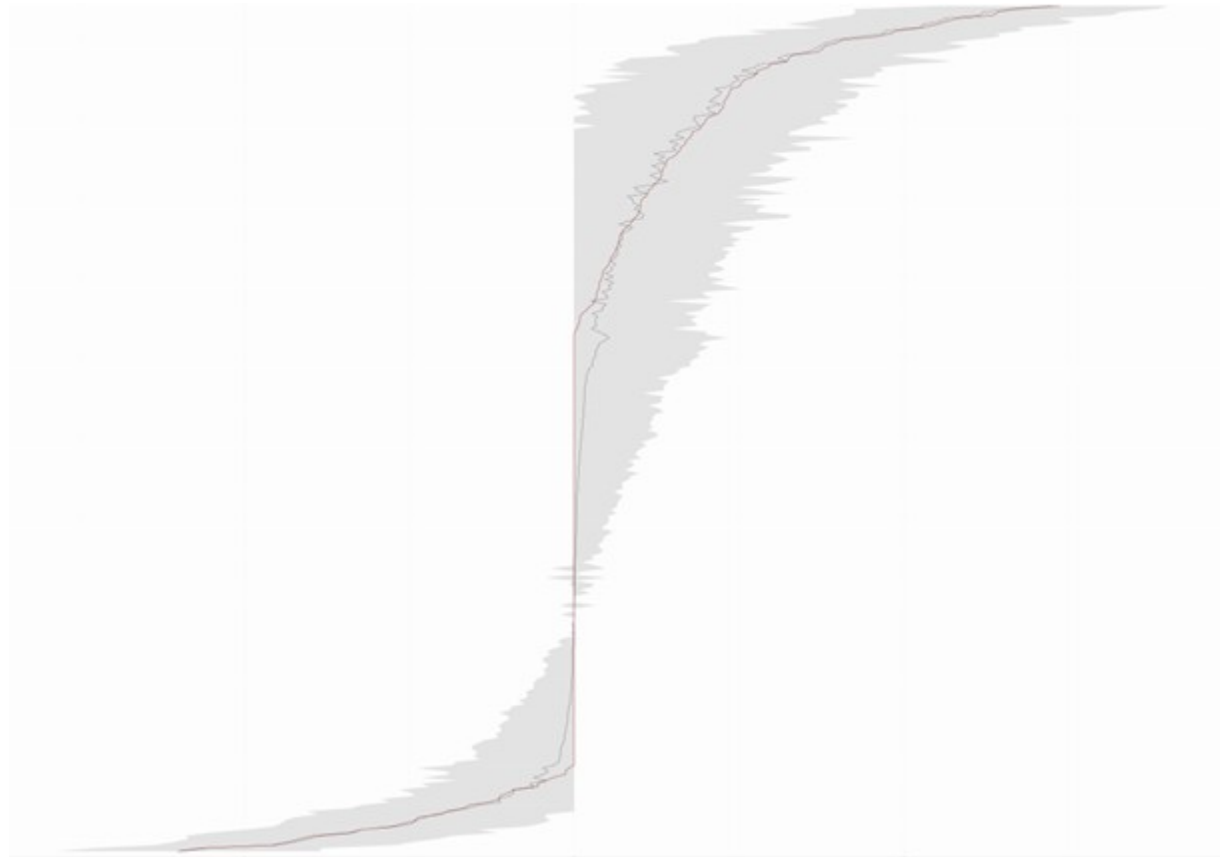
Conv	<b>-0.07997193</b>	0.00125869	0.002841678	-0.012154325	-0.021338776	0.0075
Forgive	<b>0.081865191</b>	0.036949698	0.040606246	0.03824083	0.038571429	-0.0065
Home	<b>0.080951072</b>	0.024348525	0.035046133	0.047995156	0.039306122	0.0395
Personalities	<b>0.098832776</b>	0.070655518	0.061238998	0.046265744	0.059469388	0.033
Hug	0.012694667	<b>0.227077333</b>	0.013887989	0.01866782	0.022004082	0.0375
Love	0.030125	<b>0.251762769</b>	0.014741672	0.050392659	0.043257143	0.08875
StrgConn	0.087979167	<b>0.194080645</b>	0.024065942	0.066987535	0.054853061	0.0655
OwnCommit	0.045633065	<b>0.226849462</b>	0.007422842	0.033702216	0.043987755	0.098
OnlyPart	0.046515782	<b>0.146144392</b>	0.011865489	0.040459141	0.033595918	0.0625
Sex	0.012963333	<b>0.1433</b>	0.011317814	0.086531488	0.020885714	0.00925
Fun	0.023661333	<b>0.230108</b>	0.017837382	0.082299654	0.047326531	0.0465
Outside	0.043496326	<b>0.099863727</b>	0.029258959	0.057584083	0.039195918	0.01
Happy	0.063488993	<b>0.190396264</b>	0.073981769	0.137782007	0.105840816	0.09725
Needs	0.028885333	0.157106667	0.049142375	<b>0.188734256</b>	0.069179592	0.08525
Quarrel	-0.017268667	0.005152	<b>0.044859649</b>	-0.004219377	-0.024146939	-0.0405
Angry	-0.04533956	-0.006771181	<b>-0.05882647</b>	-0.028243599	-0.032110204	-0.03375
GoingWell	0.055754	0.050270667	<b>0.106979082</b>	0.051451211	0.083146939	0.086
ComSat	0.04456008	0.033118825	<b>0.115065541</b>	0.0474609	0.060514286	0.019
TalkFight	0.068939252	0.005277704	<b>0.118870946</b>	0.006758478	0.055591837	0.03775
Separate	-0.049597594	-0.015510027	0.029108254	-0.042363322	<b>-0.064020408</b>	-0.01725
Control	-0.051686497	-0.011523396	-0.01171516	-0.002303114	-0.026518367	<b>-0.06175</b>
Stable	0.08493516	0.039850936	0.024320027	0.078739792	0.093318367	<b>0.10525</b>
Depend	0.017255376	<b>0.030304435</b>	0.004369388	0.00129481	0.007383673	0.01675
NotLoved	-0.024625333	-0.048381333	-0.01608637	<b>-0.27675294</b>	-0.036808163	-0.0245
Needed	0.029334667	0.024020667	0.010222672	<b>0.142498962</b>	0.020726531	0.01825
PartCommit	0.063618	0.071436	0.034161269	<b>0.209987543</b>	0.069612245	0.07375
PerfectPart	0.070326667	0.085024	0.028047233	<b>0.170415225</b>	0.040906122	0.04275

*Note.* The highest absolute average network loading across dimensions is highlighted in bold font.

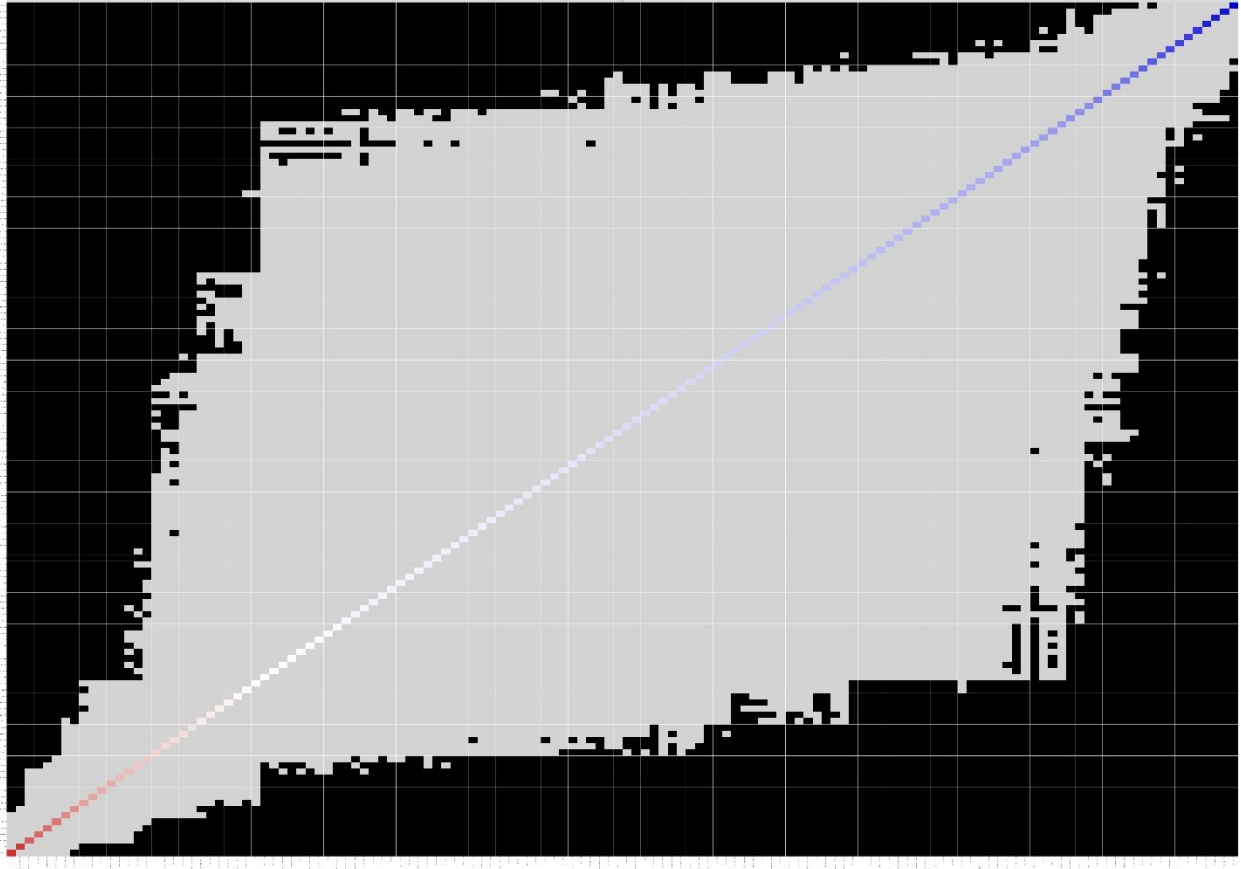


**Appendix E****Stability Analysis of the 24-Node Network****Figure E1.**

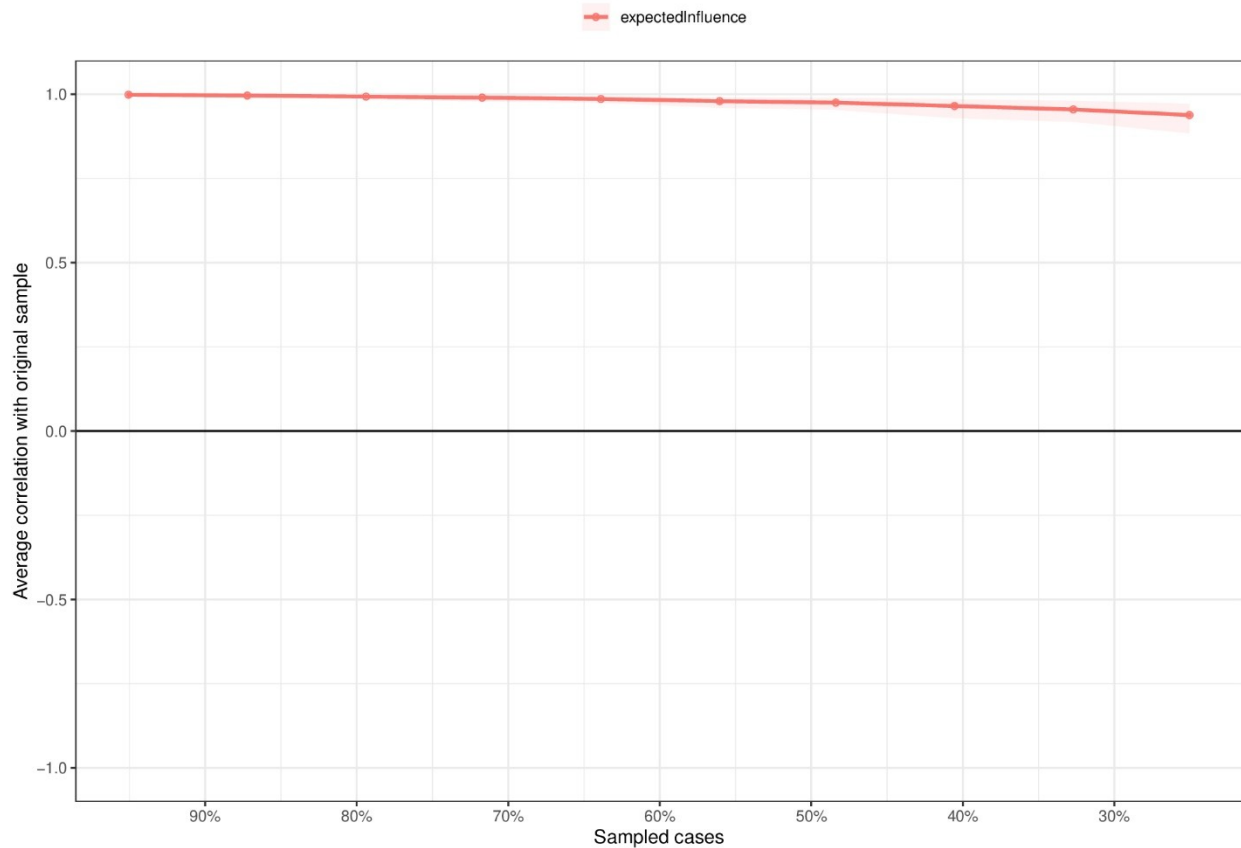
*Bootstrapped confidence intervals of estimated edge-weights*



*Note.* Each edge of the network is represented in one horizontal line and is ordered from the edge with the highest edge weight to the lowest edge weight. The grey area represents the bootstrapped confidence intervals, and the red line represents the sample values. The smaller the confidence intervals, the more accurate the edge estimates. Overlapping confidence intervals signify that edge weights are unlikely to differ from one another significantly and that interpreting the order of most edges in the network should be done carefully.

**Figure E2.***Bootstrapped difference test of estimated edge weights*

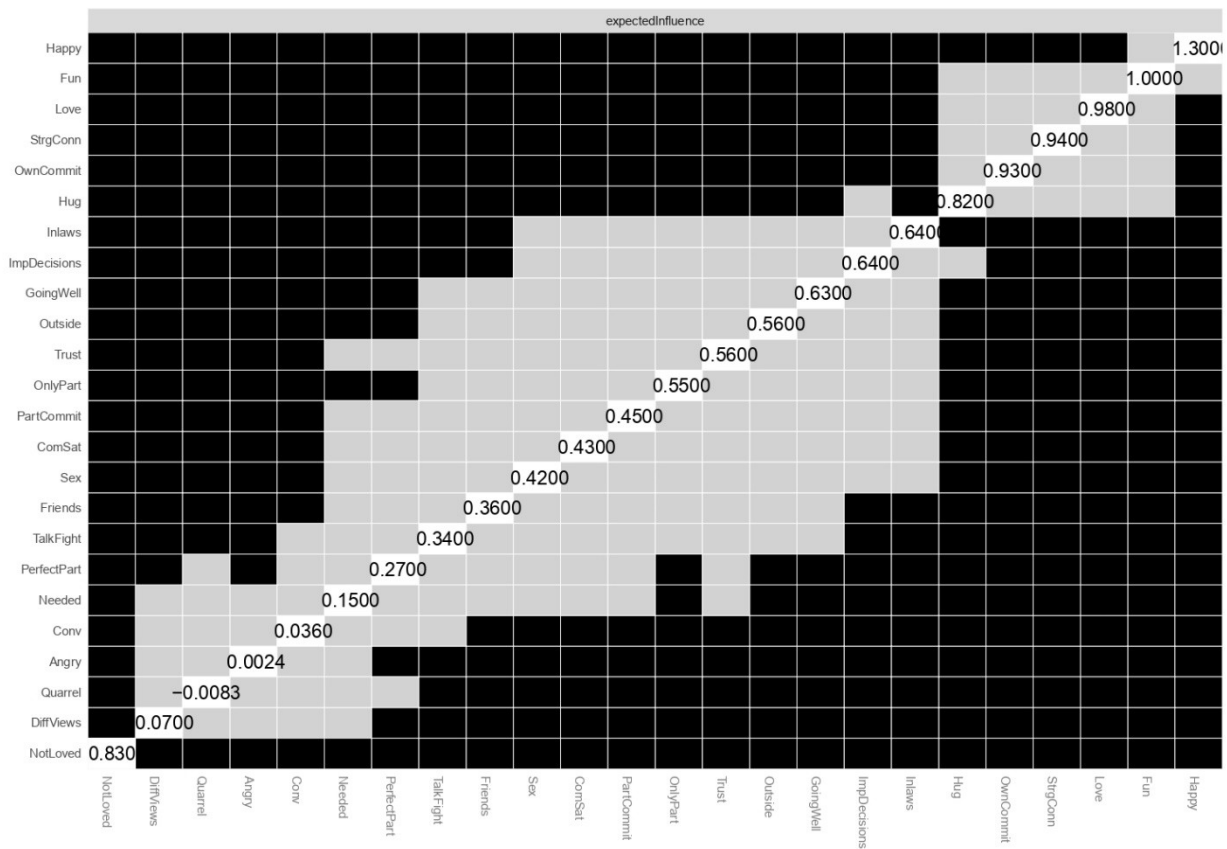
*Note.* Both axes represent all non-zero edges in the network from the lowest to the highest value. Statistical differences between two edge weights are indicated with a black box; non-statistical differences are indicated with a grey box. These results suggest that the strongest (negative and positive) edge-weights comparisons were significantly different from each other ( $p < .05$ ).

**Figure E3.***Stability of estimated expected influence*

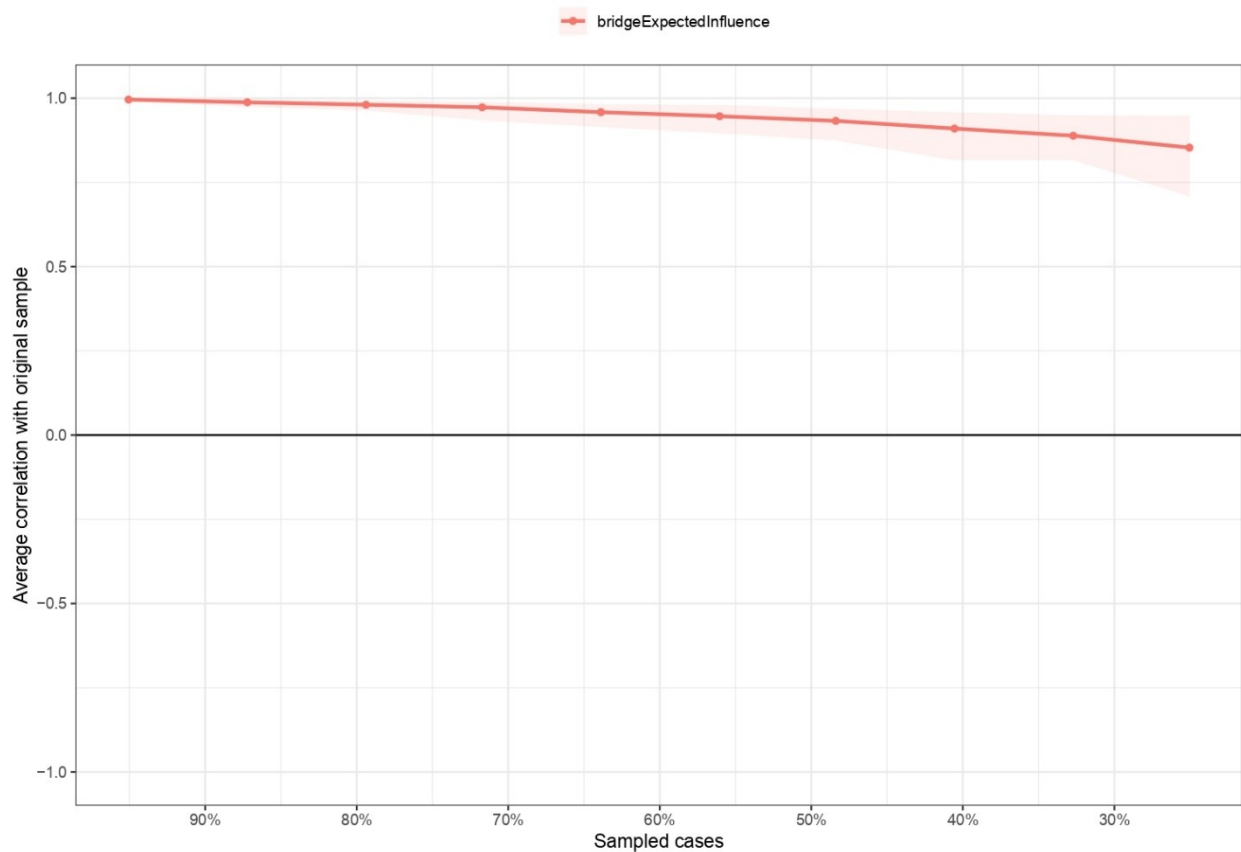
*Note.* Subset bootstrapping of the 33 items showing the correlation between the centrality values in the original network and networks estimated with fewer subjects (1,500 bootstrapped samples). The red line indicates the mean and the red areas represent the range from the 2.5th quantile to the 97.5th quantile. CS-coefficient ( $(CS(\text{cor} = 0.7) = 0.75)$ ) indicates the mean percentage of our sample that can be dropped to preserve a correlation of  $r = .7$  between our sample's centrality indices and our case-dropped bootstraps' centrality indices. CS-coefficient should not be below 0.25 and preferably above 0.5.

**Figure E4.**

*Bootstrapped difference test of estimated expected influence*



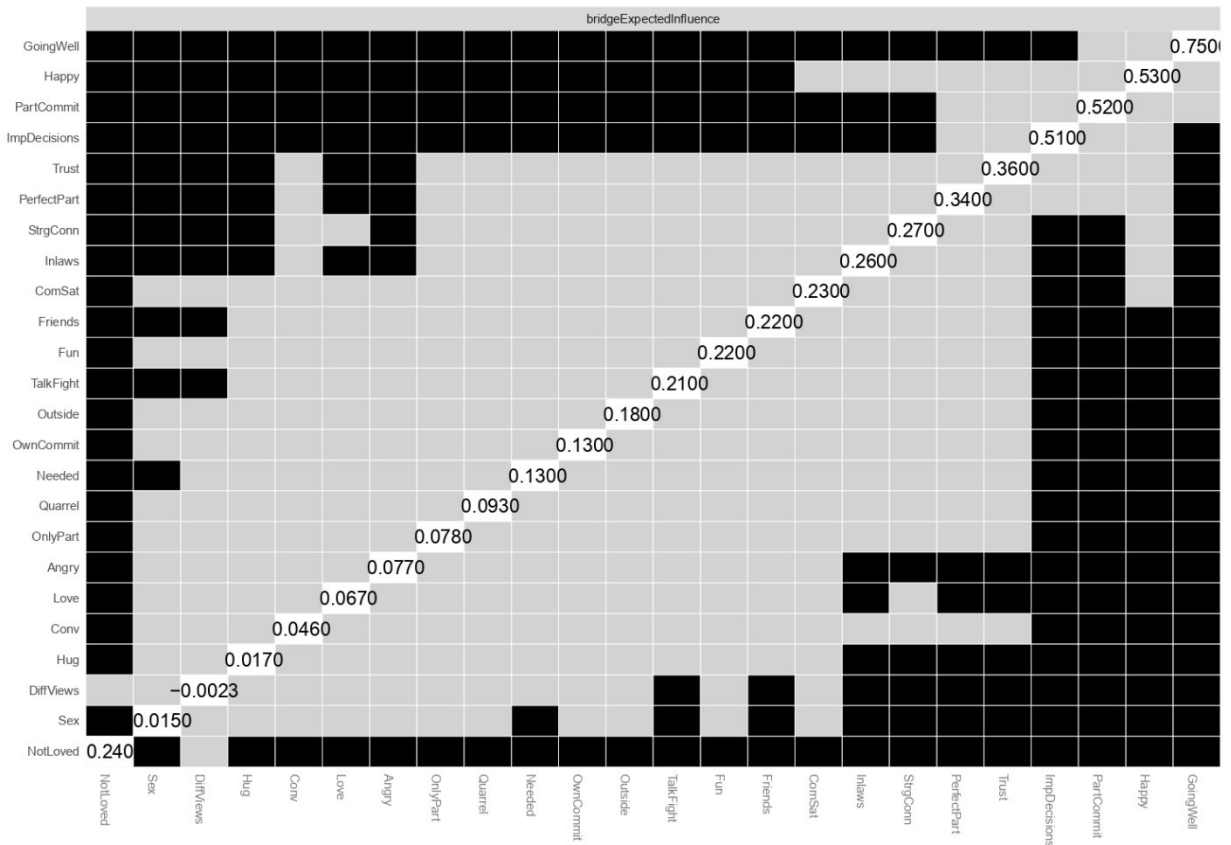
*Note.* Both axes represent all degree centrality estimates for the 33 items of the questionnaire, from the lowest to the highest value. Statistical differences between two centrality estimates are represented by a black box; non-statistical difference are represented by a grey box. These results suggest that the strongest centrality estimates comparisons are significantly different from a large majority of other nodes ( $p < .05$ ).

**Figure E5.***Stability of estimated bridge expected influence*

*Note.* Subset bootstrapping of the 33 items showing the correlation between the centrality values in the original network and networks estimated with fewer subjects (1,500 bootstrapped samples). The red line indicates the mean and the red areas represent the range from the 2.5th quantile to the 97.5th quantile. CS-coefficient ( $(CS(\text{cor} = 0.7) = 0.75)$ ) indicates the mean percentage of our sample that can be dropped to preserve a correlation of  $r = .7$  between our sample's centrality indices and our case-dropped bootstraps' centrality indices. CS-coefficient should not be below 0.25 and preferably above 0.5.

**Figure E6.**

*Bootstrapped difference test of estimated bridge expected influence*



*Note.* Both axes represent all degree centrality estimates for the 33 items of the questionnaire, from the lowest to the highest value. Statistical differences between two centrality estimates are represented by a black box; non-statistical difference are represented by a grey box. These results suggest that the strongest centrality estimates comparisons are significantly different from a large majority of other nodes ( $p < .05$ ).

## Appendix F

### Dimensions and Items Stability of the 24-node Network

**Table F1.**

*Stability statistics for the EGA network derived from 1500 bootstrapped samples*

Summary statistics for bootstrapped EGA					
N of boots	Med.dim	SE.dim	CI.dim	Lower	Upper
1500	4	0.165	0.324	3.676	4.324

*Note.* This table provides descriptive statistics for the number of bootstraps, median number of dimensions found across the bootstraps, standard error (SE.dim), and lower (Lower) and upper (Upper) bound of the 95% confidence interval of the bootstrapped dimensions.

**Table F2.**

*Dimensions likelihood for the EGA network derived from 1500 bootstrapped samples*

# of dimensions	Likelihood
3	0.010
4	0.973
5	0.017

*Note.* Proportion of times the number of dimensions was identified.

**Table F3.***Structural consistency of dimensions and average item stability*

Dimension	Structural consistency	Average item stability
1	0.695	0.971
2	0.916	0.981
3	0.984	0.986
4	0.989	0.948

*Note.* Structural consistency is the proportion of times that each empirical EGA dimension exactly replicates across the 1,500 bootstrapped samples. We also provided the average item stability in each empirical EGA dimension.



**Table F4.***Items likelihood for the EGA network derived from 1500 bootstrapped samples*

	Dimensions				
	1	2	3	4	5
Hug	<b>0.99666667</b>	0	0	0.002	0.001333333
Love	<b>0.992</b>	0.001333333	0	0	0.006666667
OwnCommit	<b>0.992</b>	0.001333333	0	0	0.006666667
StrgConn	<b>0.99066667</b>	0.002666667	0	0	0.006666667
Sex	<b>0.96533333</b>	0	0	0.033333333	0.001333333
OnlyPart	<b>0.96066667</b>	0.034	0	0.000666667	0.004666667
Fun	<b>0.958</b>	0.000666667	0.003333333	0.030666667	0.007333333
Outside	<b>0.88266667</b>	0.032666667	0.028	0.049333333	0.007333333
Happy	<b>0.82133333</b>	0.001333333	0.062666667	0.108	0.006666667
Inlaws	0.002	<b>0.992</b>	0.002	0.002	0.002
Trust	0	<b>0.99133333</b>	0.006666667	0.002	0
Friends	0	<b>0.98933333</b>	0.009333333	0.000666667	0.000666667
ImpDecision s	0.012666667	<b>0.98266667</b>	0	0.000666667	0.004
DiffViews	0	<b>0.96</b>	0.037333333	0	0.002666667
Conv	0	<b>0.934</b>	0.061333333	0.002	0.002666667
TalkFight	0	0.007333333	<b>0.99</b>	0	0.002666667
ComSat	0	0.007333333	<b>0.99</b>	0	0.002666667
Quarrel	0	0.007333333	<b>0.99</b>	0	0.002666667
Angry	0	0.007333333	<b>0.99</b>	0	0.002666667
GoingWell	0.002666667	0.008666667	<b>0.98666667</b>	0	0.002
NotLoved	0	0.000666667	0	<b>0.99933333</b>	0
PartCommit	0	0.000666667	0	<b>0.99933333</b>	0
Needed	0	0.001333333	0	<b>0.99866667</b>	0
PerfectPart	0.006666667	0.002	0	<b>0.99</b>	0.001333333

*Note.* Item stability. Each value represents the proportion of times each item is replicated in each of the empirical EGA-defined dimensions. We used the input from the EGA network derived from 1500 bootstrapped samples. The highest absolute likelihood for an item is highlighted in bold font.

**Table F5.***Average network loading for each item in each dimension over 1,500 bootstrapped samples*

	Dimensions				
	1	2	3	4	5
Hug	<b>0.241534</b>	0.00361189	0.028196776	0.004869246	0.088923077
Love	<b>0.28513</b>	0.0129666	0.023731363	0.042543029	0.136846154
OwnCommit	<b>0.24573</b>	0.054862391	0.002482203	0.036364243	0.110076923
StrgConn	<b>0.217771333</b>	0.08546493	0.030795165	0.083895931	0.116346154
Sex	<b>0.158792</b>	0.016763527	0.020524513	0.091991995	0.043884615
OnlyPart	<b>0.160424667</b>	0.066979292	0.024691068	0.022646431	0.077692308
Fun	<b>0.271071333</b>	0.017233801	0.033941572	0.069430287	0.154115385
Outside	<b>0.114592667</b>	0.055957248	0.042387508	0.055096064	0.067615385
Happy	<b>0.225624</b>	0.065796927	0.149041639	0.149306871	0.130615385
Inlaws	0.065233333	<b>0.18358517</b>	0.013011417	0.047453636	0.046269231
Trust	0.043244667	<b>0.210365397</b>	0.07565413	0.063831221	0.008307692
Friends	0.021285333	<b>0.234184369</b>	0.07819409	0.013896598	-0.010730769
ImpDecisions	0.099248667	<b>0.196077488</b>	0.049418402	0.055203469	0.060423077
DiffViews	-0.015381333	<b>-0.16370541</b>	-0.045471457	-0.018969313	0.005269231
Conv	-0.002864667	<b>-0.12640481</b>	-0.067528543	-0.013444296	0.023653846
TalkFight	0.011937333	0.092678023	<b>0.233222297</b>	0.012894596	-0.007730769
ComSat	0.072015333	0.051098196	<b>0.25022364</b>	0.080670447	0.004615385
Quarrel	0.006542667	-0.034768871	<b>-0.25767562</b>	-0.009861241	0.047961538
Angry	-0.021565333	-0.077788243	<b>-0.24886904</b>	-0.049669113	0.040115385
GoingWell	0.126906667	0.090537742	<b>0.233145064</b>	0.060608406	0.039384615
NotLoved	-0.068224667	-0.034314629	-0.062427132	<b>-0.29429219</b>	-0.017615385
PartCommit	0.100808667	0.058263861	0.044302216	<b>0.241757171</b>	0.065269231
Needed	0.028141333	0.044445558	0.018006044	<b>0.172350901</b>	0.023807692
PerfectPart	0.118398667	0.06899666	0.050858966	<b>0.182897932</b>	0.0665

*Note.* The highest absolute average network loading across dimensions is highlighted in bold font.