

# Business-oriented Analysis of a Social Network of University Students

Informative value of individual and relational data compared through group detection

Vincent Labatut<sup>1</sup> & Jean-Michel Balasque<sup>2</sup>

*Presented by Günce Orman<sup>1</sup>*

<sup>1</sup>Computer Science Department

<sup>2</sup>Business Science Department

Galatasaray University, Istanbul, Turkey



# Overview

- I. Social Network Analysis in Business Science
- II. Data Collection
- III. Methods
- IV. Results
- V. Discussion
- VI. Conclusion
- VII. Perspectives

# SNA in Business Science

## Main ideas

- **Utilitarian perspective:** benefits of SN analysis for a person or firm
  - Understand/influence personal judgment and decision making
  - Improve organization and management
- Points of interest
  - **Information propagation** (nature, efficiency, speed...)
  - Identification of **groups** of persons
- Usual method: **local focus**
  - One person (or a small group) is studied
  - Generalization to the rest of the network

# SNA in Business Science

## Limitations

- Local vs. **global** approach
  - Propagation depends on the SN **overall properties**
  - Necessity to consider **larger structures**
  - ▶ Local and global approaches are **complementary**
- Non-**systematic** approach
- Possible **causes**
  - CN tools relatively new
  - SN extraction can be costly and difficult

	Factual	Behavioral	Sentimental
Individual	--	-	+
Relational	+	++	+++

*Information gathering difficulty / cost level*

# SNA in Business Science

## Proposed Approach

- **Problematics**: is it worth using relational data?
- Focus on **group** identification
- **Systematic** and **global** study
- Systematic approach:
  - **Cluster** analysis on **individual** data
  - **Community** detection on **relational** data
  - Analysis:
    - Group **comparison**
    - **Community** membership **prediction**

# Data Collection: Galatasaray University



- Small public university (~2000 students)
- Top 5 Turkish universities
- Mainly French-speaking
- Entrance: national vs. internal exam
- Very efficient alumni network
- Strong image

# Data Collection: Survey

- Spring 2009
- 224 respondents
- Anonymous
- Themes:
  - Personal attributes
  - Social interactions
  - Purchasing behavior
  - Favorite brands/products
- **Factual**, behavioral & sentimental **individual** data
- Sentimental **relational** data



# Data Collection: Fields of Interest

- **Individual Data (Factual)**

Attribute	Value
Gender	Male vs. Female
Department	12 different departments
Class	6 different years
Entrance exam	National vs. internal
High-school category	6 different categories
High-school city	Istanbul vs. other
High-school specialization	17 possible specializations
Club membership	40 different activities

- **Relational Data (Sentimental)**

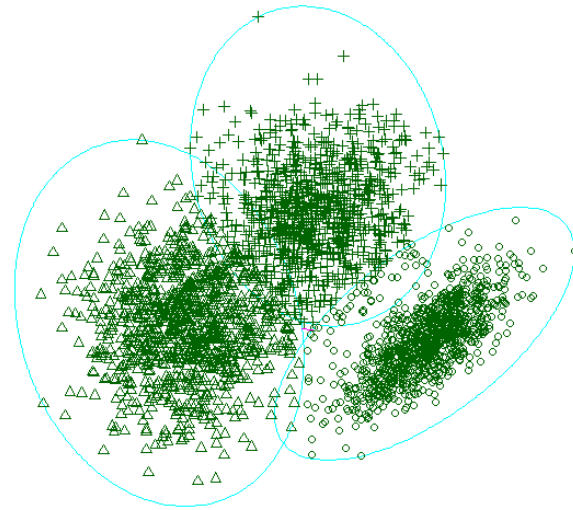
- Names of important students
- Intensity of the relationship, from  $-5$  (hatred) to  $+5$  (love)



# Methods: Groups Identification

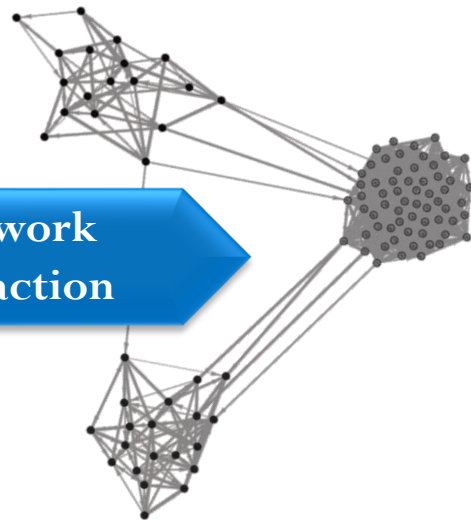
Individual  
Factual Data

Two-Step  
Clustering

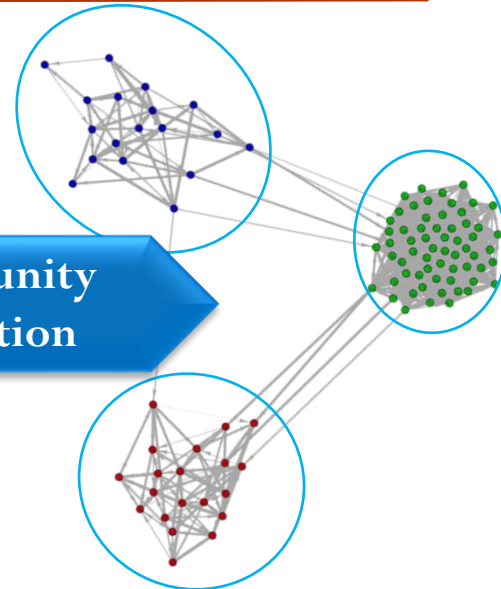


Relational  
Sentimental  
Data

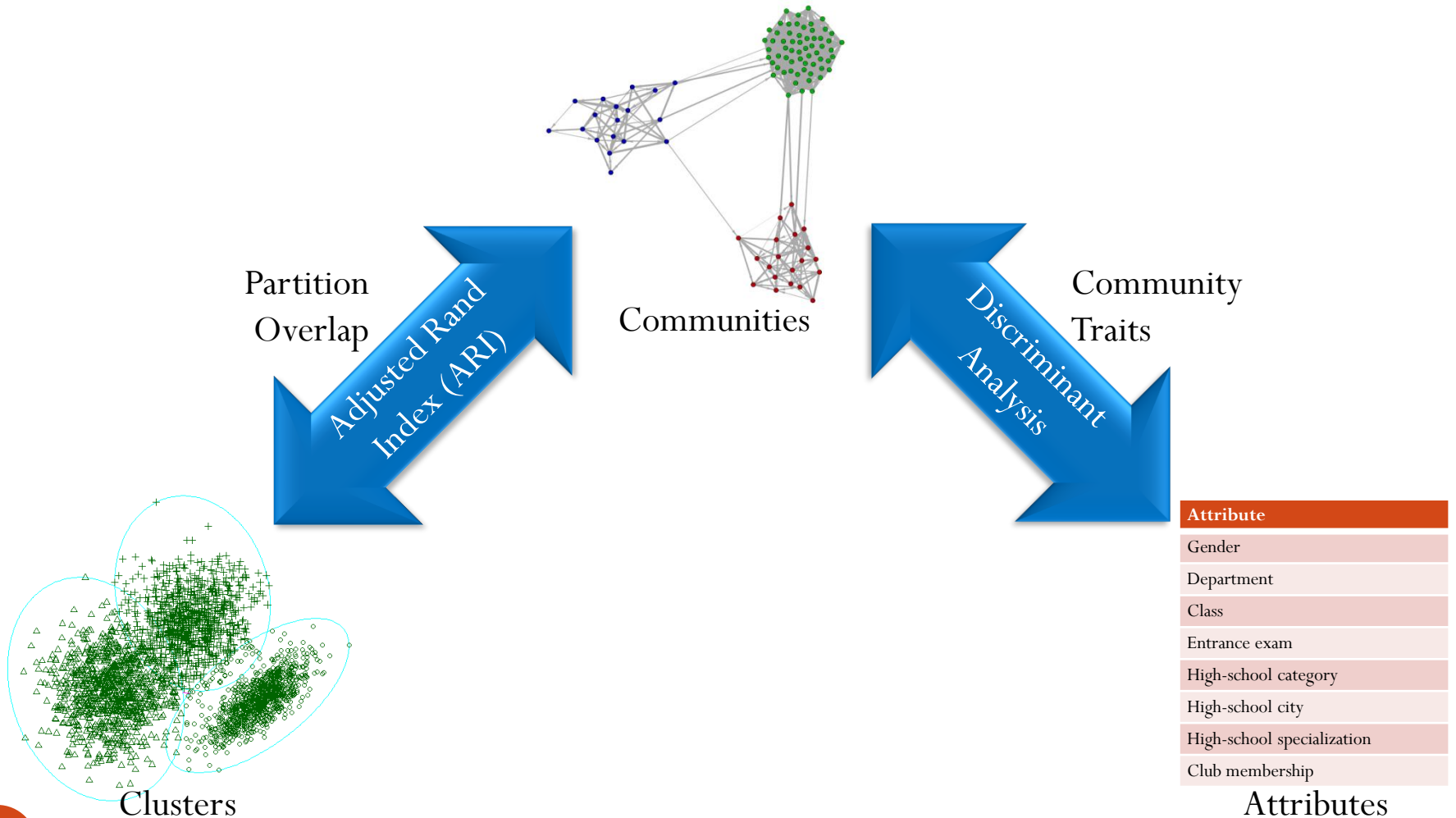
Network  
extraction



Community  
Detection



# Methods: Groups Analysis

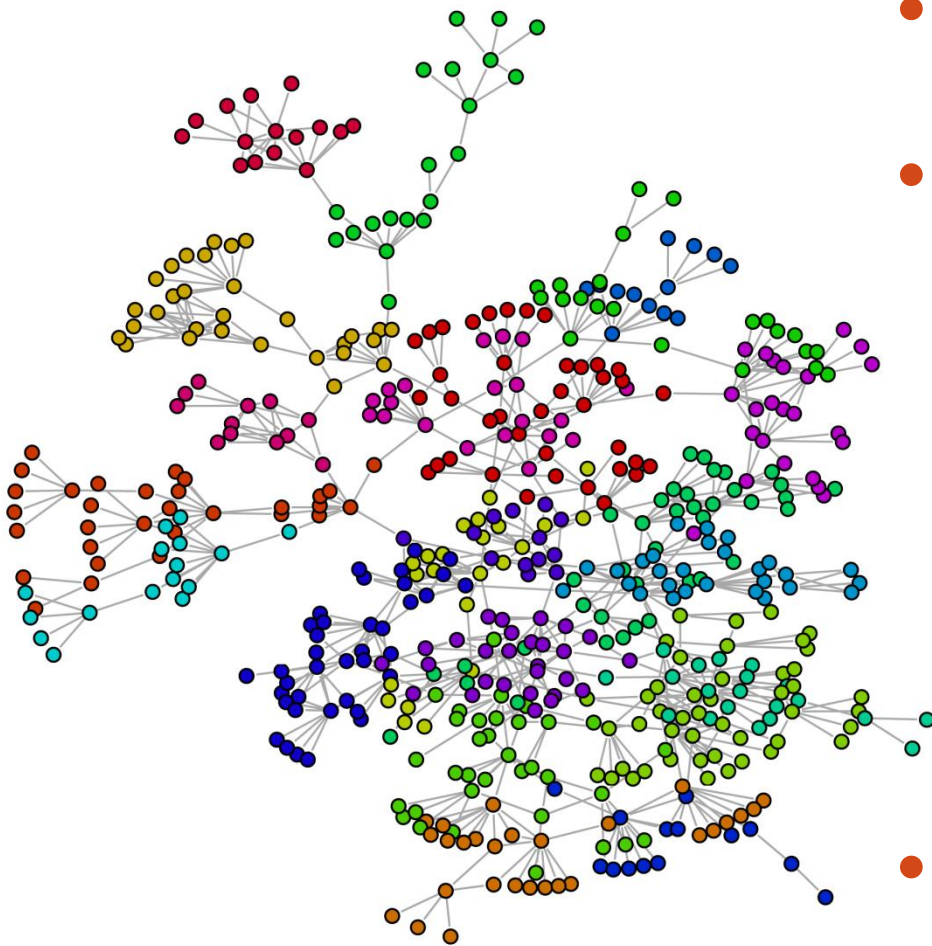


# Results:

## Clusters

- Exhaustive search
  - all combinations of factual attributes were tested
- Best separation (BIC) for 3 attributes:
  - Gender
  - Department
  - Class
- Result: 4 Clusters
- Other attributes lead to poor partitions

# Results: Communities



- Unweighted, undirected, trimmed network
- Stable results over all tested algorithms (ARI)
  - Edge-Betweenness [Girvan & Newman]: 23
  - Fast Greedy [Newman & Girvan]: 22
  - Spinglass [Reichardt & Bornholdt]: 29
  - Walktrap [Pons & Latapy]: 37
- Optimal partition ( $Q=0.88$ ): 22 communities (FG)

# Results:

## Groups Comparison

- 4 optimal clusters vs. 22 optimal communities
  - $ARI = 0.043$
- 4 optimal clusters vs. 4 forced communities
  - $ARI = 0.001$
- 22 forced clusters vs. 22 optimal communities
  - $ARI = 0.423$
  - Best separation using Gender, Department, Class, Entrance, Category, City, Specialization
- Conclusion:
  - **Low overlap** between communities and clusters
  - **Tenuous link** between communities and attributes

# Results:

## Community Prediction

- Communities are the reference class
- Exploratory approach: all combinations of attributes considered
- Correlated attributes automatically discarded
- Significantly **discriminant attributes**:
  - **Class** (year of study)
  - **Entrance** examination (national or internal)
  - University **department**
- Classification test: **37.2%** success rate

# Discussion:

## Discriminant Attributes

- Department less important than class/entrance exam
  - **Department:**
    - Expected
    - Students from the same department interact more
  - **Preparatory** years:
    - Development of cross-departmental relationships
    - Strong bounds due to context (new location, friends, language, no family)
  - **Internal** examination:
    - French-speaking students tend to stick together
    - Come from high-schools with strong identities
    - Significant social differences with other students

# Discussion:

## Non-discriminant Attributes

- Seemingly important factor are not discriminant
  - Gender
    - Young students
    - New context
    - No family
  - High-school city
    - No discrimination between students from Istanbul and the others



# Conclusion

- Groups built from a survey:
  - Clusters from individual factual data
  - Communities from relational sentimental data
- Low overlapping between the groups
  - ▶ Complementary data
- Link between the two kinds of data
  - Community can be predicted from attributes for 1/3 students

# Perspectives

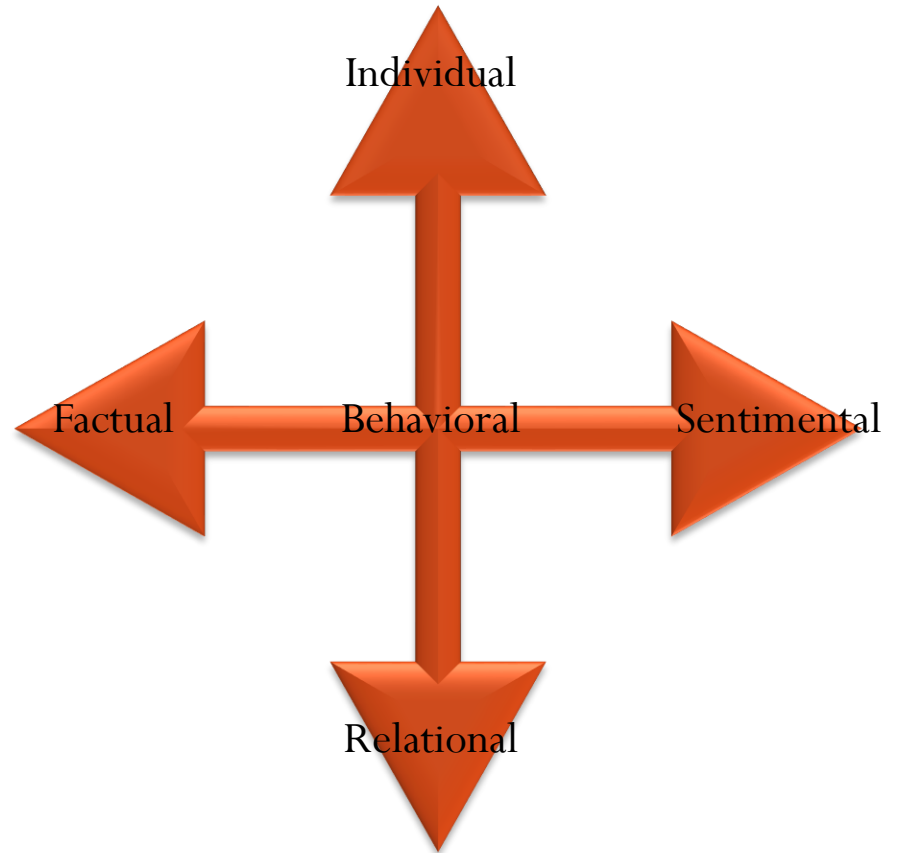
- **Limitations**
  - Small population: collaborations
  - Insufficient response rate: additional surveys
- Take advantage of the individual **sentimental** and **behavioral** data
  - ▶ Will possibly lead to different clusters
- Consider relationships **asymmetry** and **intensity**
  - ▶ Directed and weighted network
- Allow **overlapping** groups
  - ▶ One student can belong to several groups

# Questions



# Data Categories

- **Individual**: related to only one person
  - Weight, mood, income...
- **Relational**: related to several (usually 2) persons
  - Friendship, being colleagues...
- **Factual**: acknowledge information
  - Individual: gender, age...
  - Relational: emails, collaborations...
- **Behavioral**: observations
  - Individual: reaction to some situation...
  - Relational: social interactions...
- **Sentimental**: feelings, thoughts
  - Individual: tastes...
  - Relational: friendship, hate...



# Adjusted Rand Index

- **Rand Index**

- $a$ : number of pairs whose elements belong to the same group in both partitions
- $b$ : number of pairs whose elements belong to the same group in the first partition, whereas they belong to different groups in the second one
- $c$ : number of pairs whose elements belong to different groups in the first partition, whereas they belong to the same group in the second one
- $d$ : number of pairs whose elements belong to different groups in both partitions

$$RI = (a + d) / (a + b + c + d)$$

- **Adjusted Rand Index**

- $E$ : expected (chance) similarity

$$ARI = (RI - E) / (1 - E)$$