

Changes in the Article “The Problem of Action at a Distance in Networks and the Emergence of Preferential Attachment from Triadic Closure”

April 4, 2017

Dear Editor,

We are happy for the encouraging and constructive reports of the reviewers. We have addressed these in detail below. We have revised the manuscript thoroughly, taking into account most of the comments of the reviewers and have added a new section to the manuscript. We believe the manuscript is now publishable in the Journal of Interdisciplinary Methodologies and Issues in Science.

This letter describes the differences between the article “The Problem of Action at a Distance in Networks and the Emergence of Preferential Attachment from Triadic Closure” as submitted in September 2016, and the revised version of the article from March 2017.

We have added a series of empirical experiments on synthetic datasets in Section VII (Derivation). These consist in generating synthetic random networks of fixed size, interpolating between three the three extremal random graph types “Erdős–Rényi”, “Preferential attachment” and “Triangle closing”. The purpose of the experiments is to show empirically that preferential attachment emerges in the “Triangle closing” models, which is not the case for the Erdős–Rényi networks.

The description of related work was moved to the beginning of the article (Section II) to give a better overview of and motivation for the field.

Other modifications to the article are described with the comments below.

Best regards,

Jérôme KUNEGIS, Fariba KARIMI, SUN Jun

Reviewer 1

In this article, the authors describe how the mechanism of triadic closure could be the cause of the preferential attachment phenomenon. They first describe

both process (triadic closure and preferential attachment). They then describe the rationale behind their assumption, and provide a formal model illustrating its relevance. This work fits the special issues topic.

I found the paper well written and easy to read and understand. The points are well formulated, the argumentation is flowing logically in all the discussion parts, which represent most of the paper. The formal model is also clearly motivated. The idea behind the paper (triadic closure as the hidden cause of preferential attachment) makes sense from an intuitive perspective, which makes it appealing, and seems to be supported by the proposed model.

I could find a few typos:

- P.1, L.25: “On the other hand” → this expression is meant to come after “On the one hand”, in order to put in perspective two conflicting points. I could not find no explicit mention of the first hand. . .
- P.1, L.19: “if Alice likes a movie and Bob is friends with Alice” → Bob is a friend of Alice (less colloquial)
- P.4, L.29: “In recommender systems, such as that used on web sites that recommend movies to watch” → such as those used on web sites, or such as the ones used on web sites
- P.5, L.7: “The explanation for preferential attachment thus lies in hidden nodes: Nodes that make indirect connections between things, but do not appear in the modelled system.” → the system is the reality, the model is its representation. The hidden nodes therefore are necessarily present in the system, but might be hidden in the model. So: → do not appear in the model
- P.5, L.44: “two nodes in V connect 45 with a probability proportional the number of common nodes they have” → proportional to the number
- P.6, L.7: “The typical degree of nodes are significantly smaller than the number of nodes” → the typical degree is significantly smaller
- P.7, L.22: “the classic preferential attachment model lacks to explain the number of clusters” → fails to explain

We have corrected all typos.

The article is quite self-contained, and does not require too many references. However, I think the introduction lacks some bibliographic references. First, to introduce some notions defined in previous works. For instance: Who originally introduced the notion of triadic closure (mentioned P.1 L.15)? What about that of preferential attachment (P.1 L.25)? Some reference are provided later, but it would be better to give them on the first occurrence of the notion.

We have inserted the requested references at the point where they are needed.

Second, the authors sometimes state some points without supporting them with an appropriate bibliographic reference. For instance, the authors claim that “preferential attachment is true empirically, and has been verified many times in experiments”: I do not say this is incorrect, and actually I agree, but this should nevertheless be backed in some way (if it has been verified many times, this should not be too difficult).

We have added appropriate references.

I have a few remarks that could help improving the paper, and I think some questions should be answered, too. Here are my detailed comments (in no particular order), which aim at being as constructive as possible.

1) Preferential attachment.

In the introduction, triadic closure is presented as *an example* of local growth, whereas preferential attachment (PA) is implicitly presented as the *only* mechanism for global growth. I do not think this is correct, and this is probably not what the authors mean: this should be explicitly stated.

About PA itself, the authors mention in the introduction that it is the process that will lead two popular persons to become friends. But actually, only one of them needs to be popular for this to happen, according to PA (if a node has a high degree, it has a high probability to get attached to other nodes, whatever their degree). And this remark is made later by the authors themselves, when discussing recommender systems (P.4, section Explanations). Maybe the authors meant both persons are looking for friends in their example (by opposition to only one node looking for new connections in the BA model), but this is not clear. So this should be clarified in their introductory examples.

We have clarified the relationship between the two concepts.

2) Synonyms.

The article would be clearer if the key concepts were always named using the same expressions. The use of synonyms can lead the reader to think these lexically different expression point at different semantically different concepts. Here, the authors use most of the time “triangle closing”, but also sometimes “triadic closure” (including the title). They also seem to use “primitive”, “basic” and “fundamental” as synonyms when referring to phenomena or rules. To my opinion, using a more uniform lexicon would increase the article clarity.

We have added a note about the terminology.

On the same note, the notion of “fundamental phenomenon” is central to this article, but is clearly defined only in the discussion: I think this part should be moved in the introduction instead.

A clarification has been added to the introduction.

3) Is really everything a network?

I found the discussion in the Network section very interesting. And I also think the authors are quite provocative by stating that “everything is a network” (P.2, L.27). First, what they explain right before this statement is that networks

are used in many, if not all, scientific fields (I would add they are now used even in non-scientific fields, such as literature). But being the most popular modeling tool does not mean being the only one, nor does it mean that everything can be represented as a network.

Second, as illustrated by the authors themselves in the rest of this section, a lot of data could be qualified of “individual”, in the sense they characterize some objects independently from the other objects belonging to the same dataset. For instance, in a social group, attributes such as: height, age, job, gender, etc. By opposition, graphs are useful to represent “relational” data, describing the dependencies between objects: friendship relation, family, co-worker, and so on. It is true that individual information can be encoded in a graph under the form of nodal attributes, but this is relevant only if the dataset also contains relational data. For these reasons, I disagree with the authors’ statement (everything is a network), and I do not think the arguments they expose in the article are sufficient to make their point. Everything could be modeled through a network, but 1) this is true of any modeling tool, and 2) this does not mean networks are always the best modeling tool.

We have toned down the universal statements about networks; they do not impact the main argument.

4) Analogy.

In the third paragraph of section Preferential attachment, the authors state that preferential attachment is not a fundamental mechanism in the creation of ties in networks, because it relies on the ability to perform actions at a distance (“preferential attachment *cannot be* a fundamental driving force for tie creation” – emphasis added by me). The point is clearly explained, and intuitively it makes sense. However it is based on an analogy between a specific system, that of gravitational physics, and a modeling tool, graphs. To my opinion, this is borderline fallacious, and here is why. This modeling tool is generic, it can be used to represent almost any system (as stated by the authors earlier). There is no a priori reason to think all existing systems behave like gravitational physics, even when focusing on the specific point of distant influences. So, the inadequacy of distant influences does not necessarily translates into any considered systems, which in turn invalidates the analogy. The authors’ discussion does therefore not constitute a proof of the inadequacy of PA in graphs.

To use another analogy: physics has been used several times as a model in social science, and this was not always relevant (e.g. the first econometrical models straightforwardly adapted from thermodynamics). To my opinion, the authors’ analogy should be presented not as a hard justification that PA is irrelevant, but rather as a way to justify the authors’ *assumption* that PA is not a fundamental generative mechanism.

We have reformulated the key passage taken into account this limitation.

5) Model.

I find the graph-based description of the model very confusing. G is defined as a bipartite graph, each link connecting one node from set V to one from set W .

But at the same time “two nodes in V connect with a probability proportional the number of common nodes” Then “Edge between nodes in V will not be considered” and later “Let $u, v \in V$ be two nodes of the network [...] the probability p that u and v are connected can be derived ...”. This looks rather contradictory to me.

Maybe the authors should define two graphs: the first representing the actual system, with only one type of nodes, and links between whatever nodes, and the second representing what is observed of the system, containing two types of nodes (hidden vs. observed) and only links between nodes of different types (i.e. a bipartite graph). Otherwise, it is hard to precisely understand the model: some links are in the graph, but “ignored”... For instance, does the degree used in the rest of the section take these ignored links into account, or not? This really needs to be clarified, it is a strong prerequisite to publication.

The notation has been updated to reflect the used structures accurately.

Also, in the derivation of the model properties, it would be better if the authors would provide bibliographic references for the mathematical properties they use (for the limit, and so on). Also, the authors should number their equations to ease later reference and discussion (I would have liked to use such numbers to precisely refer to the equations concerned by my previous remark).

We have kept the unnumbered equations as per the journal's style guide.

6) No validation.

The authors make a promising assumption and present a nice model to illustrate how it could be put into practice. However there is no experimental validation of this model: authors generally perform numerical simulations to show their closed forms are corroborated in practice, e.g. Barabasi & Albert in their seminal paper on PA. This is not expansive in terms of work, and I think it would nicely complete the article. More interestingly, this would allow studying other topological properties of the produced networks: average distance? Transitivity? Community structure? Etc.

We have added the Section “Experiments” to fill this gap, as described above.

7) Perspectives.

My last point is a question (or rather a series of related questions): Would the authors suppose that, like PA, homophily is, a consequence of triadic closure? If yes, then could the fact some networks are weakly hemophilic, or even heterophilic, be also due to the presence of too many hidden nodes?

More generally, do they think triadic closure, as a fundamental phenomenon, could explain other widely observed topological properties? (e.g. degree correlation, transitivity, community structure...) If the authors can comment on these points, this could be used to enrich their discussion.

We have added these issues to the discussion.

For publication I think the authors should at least:

- Correct typos
- Correct and/or answer the questions regarding the minor points (points 1-4 and 7)
- Clarify the model description (point 5)

I do not set experimental validation (point 6) as a prerequisite for validation, but I think it would nicely improve the paper.

Reviewer 2

The claim of this paper is an original concept that links the global phenomenon of preferential attachment to triadic closure. The idea, that totally makes sense to me, is that in a few systems, preferential attachment seems to be independent from triadic closure (friends recommendation) because the triad that is closed is actually invisible from the system. Authors give a lot of illustrative examples : if you don't have any friends yet on facebook, but you create a connection with someone, it is probably true you know this "someone" in real life that is an invisible part of the system. Or it may have been recommended by the recommender system which can also be considered by an invisible node we do not model in our system. Basically, the idea is that when two nodes connect, it is not a pure random mechanism configured by node degrees, it is probably because of a recommendation issued from the system we model, or of nodes hidden from the system.

Authors then model a bipartite graph with visible and hidden nodes, and show that this idea seems realistic and allow to model and explain preferential attachment.

Overall decision: Accept

Reviewer 3

This article was submitted to a special issue on "Graphs & Social Systems". The authors argue that the preferential attachment rule for the evolution of networks can be derived from the triangle closing rule, and consequently triangle closing is revealed as the basic rule for the evolution of networks. Most of their examples and major arguments are from the social sciences area. Further, social Systems can be represented by a network, which is a graph. Thus the article fits the topic of the special issue.

The article is well written and read easily. However, to improve the quality, I have the following remarks/suggestions.

- I suggest a different title that better reflects the objective of the paper.
- Page 1, line 25: there is "On the other hand, ...", however I could not find the first idea opposed to this one.

- Page 2, lines 5 and 6: avoid the use of questions to discuss your idea. Using this particular question you raise a claim that the reader has no reason to consider. If you consider important to clarify this claim to some reader that might have the doubt, do not use a question.
- Page 3, lines 23 to 28: use “” or somehow to highlight the quote
- Page 3, lines 29 to 32: again, do not use questions to expose and discuss your ideas. My suggestion is to rewrite these four lines. The use of these questions confuses the reader. A direct statement is more enlightening.
- Page 4, line 29: replace “such as that used” for “such as those used”
- Page 5, lines 20 and 21, rewrite this text to use direct statements and remove the questions.
- Page 5, line 45: replace “proportional the number” for “proportional to the number”
- Page 7, first line: replace “and using the limit” for “and using again the limit” and remove the “again” from the end.
- Page 7, line 20: replace “the an object/person” for ?.
- Page 7, line 23: “been make” for “been made”.
- Finally Section “Related Work” should be removed to the beginning of the paper, to the introduction section, as it gives some insight about the theme.

The errors pointed out in the previous list were corrected.

The article is somehow an opinion article. Even though, some references are missing either to support the used definitions or to support considered assumptions. For instance, the introduction has no references at all. The used basic terms (triangle closing, fundamental mechanism, preferential attachment, ..., falling factorial) could have a reference for a work where its use and definition can be found. Besides that, an informal definition could be given the first time they appear in the text. Further, several assumptions used to stand your opinion could have a reference of its use.

Certain points have been clarified, and adequate citations have been added.

There are some points needing clarification.

- Is it triangle closing or triadic closure? You interchange their use. You should make it clear that it is the same mechanism.
- Page 2, around line 10: what is a fundamental mechanism? You want to establish that the triangle closing mechanism is a fundamental one? Where is it defined?

- I suggest you to rewrite the section “Networks”.
- Page 2, line 17: “everything is a network”. Why is this true? I do not agree! We can associate a network to many, many systems. We can find networks in many fields. Networks are important and play an important role in many fields. I agree with all of that. We can find networks in every scientific fields. But considering “everything is a network” is too much. And at the same time it is too restrictive.
- Page 2, line 27: “To find an answer, it is instructive to consider the field of machine learning”, Why?? Why are you restricting to this field to find such a common structure. Formalism and abstraction is needed to be able to make such comprehensive claims.
- Page 3, line 2: “not all do”: do what? please clarify.
- Page 3, lines 2 to 3, “bag of words” please add some references here.
- Page 3, lines 15 to 16, please add some references.
- Page 3, lines 23 to 16, please add some references here.
- Page 3, line 33 and following, I could not see the point of this example. In my opinion this does not stand as a justification for the social networks and should be removed.
- Page 5, line 21: “Imagine a network, ... ” → “Consider a network, ...”
- Page 5, line 35 and following: the “model” you refer to is the well known graph, I suggest you to define the graph
- Page 6, line 43: G , as stated, is a graph
- Page 6, line 43: G is a bipartite graph, make it clear and explain it better.
- Page 6, line 28, expression 7: you missed the “1-” in the last term of the expression
- Page 6, line 29: replace “with the limit” for “at the limit when n goes to infinity we may assume that”
- Page 7, line 2: it is not clear how you obtained such value for p . Please explain it better.
- Page 7, line 10 (at top): “is proportional to both u and v ” → “is proportional to both $d(u)$ and $d(v)$ ”
- Page 7, line 15: only on page 7 you give an informal definition of the triadic closure!!

- Page 7, line 27: be more precise and rewrite the sentence. “Hence, the scale free of networks and the abundance of triangles ...” → “Hence, the scale free of networks and the abundance of triangles in social networks ...”. There are networks not satisfying these properties and there are networks where, for instance, triangles are not desired.
- Page 7/8, line 40 (first line): network science?????
- Page 8, line 45: “network models” → “growing network models”
- Page 8, line 45: “which triangle closing” ?? Are there more than one triangle closing mechanism??? Explain.

We have made changes accordingly. In particular, the terms “triangle closing” / “triadic closure have been clarified, the section “Networks” has been updated, and the hyperbole and everything being a network has been toned down.

Overall decision: Publish with major modifications

Even if it is easy to convince someone that the preferential attachment is a fundamental network growth mechanism the authors only present soft arguments. The authors present a graph that attempts to describe the situation however they do not give an evidence of its veracity. As a stronger argument a set of computational experiences supporting their conclusion could be presented. At least with some computational experience the the conclusion would be more supportive and the authors would have a stronger argument.

I’m pretty skeptical about the arguments. Some stronger argument is missing.

We have expanded the discussion, as well as added an experimental section.