The Life and Death of Social Networks: A network formation model for opinion dynamics

Mikkel Werling[1]

[a]Aarhus University; [b]University of California, Merced

This manuscript was compiled on February 3, 2022

One of the most robust effects governing social life is homophily and triadic closure. We like people similar to ourselves, and we find new connections through our existing connections. These two effects can combine to self-reinforcing processes of shaping both social networks of social agents as well as their opinions. Here I investigate a stylized agent-based model of opinion dynamics. I find that the model can generate realistic social networks while simultaneously producing realistic opinion distributions.

opinion dynamics | computational social science | agent-based modeling

This template is provided to help authors submitting working papers for the Interdisciplinary Laboratory of Computational Social Science (iLCSS) at UMD. The template was developed based on the PNAS template. Examples of commonly used commands and features are listed below to help you get started with .

Your introduction goes here!Please start your introduction without including the word “Introduction” as a section heading (except for math articles in the Physical Sciences section); this heading is implied in the first paragraphs.

Sections: Theory, Methods, Results, Discussion...

1. Theory

One of the most robuts mechanisms governing social networks is the tendency of similarity to breed connection. This effect is often characterized as homophily. The effect of homophily results in neighbourhoods within social networks, which are homogeneous with regard to sociodemographic, behavioral and political characteristics. This has dramatic implications as one’s social world defines what pieces of information, opinions and interactions that an individual have. The most clear example is echo chambers in social media networks. In echo chambers, homophily can lead to a distortion of what kind of information is presented to each individual.

It is examples like those from echo chambers that clearly illustrates that the principles of homophily not only influences who we make connections with, but also how we think. Several studies have pointed to the fact that one’s peer group is an important influence on one’s behavior. For instance, shared political beliefs predict more behavioral involvement with one another.

Homophily and its effects become especially potent as they can create self-reinforcing mechanisms when combined with tie-formation principles such as triadic closure. Triadic closure refers to tie formation between "friends of friends" in a network. For instance, if $A \leftrightarrow B$ and $B \leftrightarrow C$, then this small system would achieve triadic closure by forming the edge $C \leftrightarrow A$. When new ties are found via triadic closure,

To see why homophily and triadic closure can create self-reinforcing homophily effects, notice the following. Due to homophily, the probability of A being similar to B will be higher than When they find new connections via triadic closure i.e. "friends of friends", In such cases, any social agent will have a propensity to like agents which are more similar to themselves. Moreover, they find new connections primarily by finding them via "friends of their friends".

Homophily and its effects become especially potent when social agents can be influenced by their peers.

A. Network Formation.

A.1. Social Networks. Social networks differ in key ways to more idealized networks. Here I outline the most important characteristics, which have been found across many different domains and networks.

The first of which is that the average path lengh in social networks tends to be small. This is best exemplified by the idea of "six degrees of separation", which refers to the notion that you are never more than five intermediaries away from any other person on the planet. Average path length is as the average of all shortest paths between nodes. CITE(Meeting strangers and friends) Recently, this notion has been studied empirically on social media networks, including 721 million people. They found that the average shortest path length was 4.74, corresponding to closer to four degrees of separation CITE(four degrees of separation). This feature is inpart also the claim to fame of small-world networks. Simple lattice networks can be modified with random rewiring of connections, making some connections act as shortcuts across the network. CITE(small world)

The second characteristic is the fact that social networks tend exhibit clustering behavior. Small hubs of the network are well-connected locally, while often having limited connections to other hubs of the network. The level of clustering is often quantified with a clustering coefficient. Here, different

Significance Statement

Authors must submit a 120-word maximum statement about the significance of their research paper written at a level understandable to an undergraduate educated scientist outside their field of speciality. The primary goal of the Significance Statement is to explain the relevance of the work in broad context to a broad readership. The Significance Statement appears in the paper itself and is required for all research papers.

Please provide details of author contributions here.

Please declare any conflict of interest here.

[1] A.O.(Author One) and A.T. (Author Two) contributed equally to this work (remove if not applicable).

[2] To whom correspondence should be addressed. E-mail: author.twoemail.com

coefficients can be used, but arguably the most common is the average clustering coefficient proposed by Watts and Strogatz in the seminal paper on small-world networks CITE(small world). The basic intuition is that with higher clustering coefficients, an agent's connections tend to be connected to each other. This is the reason why the literature often includes triangles and closed triangles in their terminology. A triangle is a set of three nodes, where either two of them are connected or three of them are connected. When all three are connected, the resulting set is a closed triangle. Clustering can also be thought of as the ratio between all closed triangles of the graph and all the triangles of the graph.

The third characteristic is that the degree distributions of networks tend to have "fat tails". In this case, the degree distribution refers to the probability distribution of degrees in a network. What is meant when degree distributions are described as having fat tails, is essentially two attributes. The distributions tend to have relatively few nodes with the average degree of the distribution, but instead have many low and high degree nodes. These degree distributions are also referred to as being "scale-free" or following a power-law distribution. In other words, most nodes have relatively few connections, but a few of them have disproportionally many connections compared to all the other nodes.

The fourth characteristic is that high degree nodes tend to connect with other high degree nodes. This is also referred to as nodes having positive assortativity.

Idealised networks such as the small world network and the scale-free network can generate a subset of these characteristics. However, none of these models can generate networks which exhibit all features simultaneously. When evaluating good candidate models for how social networks are generated, our candidate models should be able to generate networks which exhibit all characteristics at the same time using one underlying algorithm.

A.2. Candidate Models. One model which generates all the characteristics using a simple network generation model is the model proposed by CITE(Jackson and Rogers). The model is a simple network formation model, which creates a network by constantly adding new nodes. The primary mechanism of interest for this paper is the mechanism used for generating new connections. In their model, links are primarily formed via "triadic closure" and randomness. Triadic closure refers to the idea of closing triangles. In other words, new connections are formed primarily by searching through "friends of friends". When they are not, they are added randomly between agents. Let us first consider why this mechanism of tie formation will likely generate the typical characteristics of social networks.

In terms of the average path length, the random component of tie formation decreases the average path length considerably. This will create the aforementioned shortcuts in the network as seen in small world networks CITE(small world). Clustering is almost guaranteed in this model, as the defining feature of the model is closing triangles. The degree distribution will also be quite similar to social networks and exhibit scale-free tendencies. To see this, notice that larger degrees will have a higher probability of gaining a new agent via triadic closure. The growth of an agent's degree is proportional to their existing degree. This relation will lead to the large disparities in degree observed in scale free degree distributions. Finally, positive assortativity is likely as older nodes in the network will tend to have larger degrees as well. Older nodes will therefore also have more frequent opportunities to connect to each other, which will lead to positive assortativity.

This model was expanded upon by Ilany and Acay (CITE) which contributed in two notable ways. Their model was a model of social inheritance of different social systems in animals. They showed that their simple model could accurately capture the complex social structures of many different types of animals. This adds credence to the idea that triadic closure might be a common tie formation strategy throughout biology. The model also made an important alteration to the model by Jackson and Rogers (CITE) by making the network a finite size. The original model by Jackson and Rogers, the network grows indefinately. Ilany and Acay modified the network to function more closely to biological networks. This is done by having a probability for agents to dissapear from the network, which stabilizes the size.

Beyond being able to generate the attributes of social networks, the model proposed by Jackson and Rogers (CITE) is in accordance with existing theory. Triadic closure is the most common structural constraint of real life networks. In minimal dynamic social networks, traidic closure should be considered the main mechanism for tie formation. In more general terms, it has been found that the likelihood of forming a new tie is a monotonically decreasing function of distance. For instance, forming new links to people further than 4 degrees of separation from you is 2.500 times less likely than forming new links via triadic closure CITE(Origins of Homophily). Moreover, triadic closure represents a simple and natural mechanism - it lines up with common sense intuitions of how social networks develop.

A.3. The problems with current models. As mentioned previously, many of the idealized models for generating social networks can only partly generate the characteristics of social networks. Arguably, this is a symptom of a deeper problem in the field which is a loose connection between models and empirical data. An example of this is the fact that even though we refer to the distributions of social networks as "scale free", scale free networks don't match observed social networks. Few models even include data, or calculates how well their model approximates the system they are modeling. Often such measures of fit are eye-balled and not rigorously defined.

We therefore need better integration between models and data. This is true both for model creation but also true for model evaluation. As is always the case, a model has to simplify the world and make crude assumptions. A good model is a useful model with useful assumptions. The results from a model critically hinges upon these assumptions. Evaluating the assumptions of a model is therefore a crucial step to take before jumping to conclusion. This is especially noteworthy as very few of the current models of social influence are dynamic networks. This assumes that no new relations are created, and that no relations are deleted. This might not seem consequential, but it is. The model will assume that relations are kept between people even when they vermently disagree about everything. This fact in itself can account for much of the signal reported in classical models in the literature.

B. Social Influence. Including some of the basic literature (Axelrod)

B.1. Homophily. One of the most persistent facts regarding social networks is that similarity breeds connection. This tendency is known as homophily where individuals seem to be connected to individuals like themselves. This is not only true for humans, but is a pervasive fact for tie formation in numerous biological systems. From zebras to dolphins, homophily predicts which ties will companionships will form. In primates, similarity even predicts the quality of the tie you have with other people. The more you have in common, the more likely that person is a close friend. This is also true for humans. We exhibit high levels of homophily in tie formation. Personal networks in humans are homogeneous across sociodemographic, behavioral and intrapersonal characteristics. Characteristics like age, sex, race, education level, intelligence, attitude and aspirations have all been shown to exhibit high degrees of homophily. This also exemplifies what is meant when it is said that homophily is an extremely robust finding. It permeates almost every part of social life. Studies regarding the homophily of race in schools found that in middleschool, only 10% of the expected cross race friendships were observed. A similar example comes from religion. A study concerning the social networks of Jews found that 80% of their friendsships were with other Jews. In addition, 80% of marriages were with other Jews. At the time, Jews consisted of only 2% of the population.

The underlying mechanism of what causes homophily is not completely clear. The most plausible explanations are individualistic and structuralistic explanations. The individualistic explanation to this effect is the psychological claim that people will tend to prefer people who are similar to themselves. It is in other words the psychological preferences of the individual. The structuralistic explanation to this effect is that structure of the environment will make certain choices more or less available to the individual. If most of the people in your neighborhood are similar to you, chances are that you will form ties to people similar to you, regardless of your psychological preference. These two different causes of homophily are important to distinguish, and I will follow the definitions proposed by Kossinets and Watts. I will define the effect of the individual's psychological preference as choice homophily and the structuralistic effect as induced homophily. The literature suggests that neither induced nor choice homophily is enough in themselves to explain the patterns of similarity in social communities. In the study by Kossinets and Watts, highly similar pairs were 50 times more likely to form a tie than dissimilar pairs, and 13 times more likely than average similarity pairs. However, when controlling for structural constrains (i.e. induced homophily), both numbers drop significantly (4 times as likely than dissimilar pairs, 2.5 times more likely than average similarity pairs). In other words, induced homophily is responsible for much of the observed homophily, but there is also a strong effect of choice homophily regardless.

Many of the personal characteristics that show the strongest homophilic effect (i.e. race, background, religion) is always or often inherited by parents of the individual. This fact also suggests that even small amounts initial choice homophily can lead to extreme levels of induced homophily by amplifying the effect over the course of generations. More similar individuals are more likely to interact and interaction between them is likely to make them even more similar through social influence. This is also in line with findings suggesting that peer groups are an important source of influence for the behavior of people. This points to the possibility that powerful feedback loops between homophily and social influence are likely to be an underlying mechanism shaping our opinions and our social networks.

B.2. Xenophobia.

B.3. Shaping opinions. Introduce the evidence from psychology and computational literature to show why the assumptions in the model make sense

B.4. Models of Social Influence. Report the evolution of models and where to place this model in all of the literature

C. A network formation model for social influence. Explain the importance of making both a network formation and opinion dynamics model in one go

2. Methods

A. Model specification. The model is a network formation algorithm, relying on the opinion of the agents to determine how tie formation and tie dissolution happens. These opinions in turn rely on the opinion of other agents via social influence. The network is therefore truely dynamic over time in regards to both its connections and the opinions of the agents.

For clarity, we will here give a more rigorous mathematical formulation of the model. Let G be the graph of the network and t be the timestep of the model. We denote G_t to refer to the state of graph, G, at timestep, t. Let N be current number of nodes in G and N' be the target number of nodes. By target number of nodes, I mean the number of nodes we want the final network to contain. As the model is a network formation model, the number of nodes N for G_0 will always be 0. We then model the probability of agent genesis, $P(G)$ and agent death $P(D)$ simply as:

$$P(D) = \frac{N}{2N'}$$
$$P(G) = 1 - P(D)$$

At each t, a node is either added or deleted from G according to the probabilities specified by $P(G)$ and $P(D)$. By modeling node genesis and death this way, we ensure that for a large enough t, $N \approx N'$. When an agent is deleted from G, it is done so by sampling a node from G randomly. All nodes are initialized with an "opinion"-value, O. This value is drawn from uniform distribution, with a lower limit of -1 and upper limit of 1:

$$O \sim \mathcal{U}(-1, 1)$$

This ensures that agents are initialized without any bias in opinions. Moreover, this conceptualization of opinions offers logical interpretations, as 0 will be the neutral middle of the opinion space.

In the special case of $N = 0$, a new node is added without adding extra connections. For $N \leq 1$, a new node is added and connected randomly to another node in G. With probability

$P(E)$, the new node will add another connection to a randomly sampled neighbor's neighbor. With probability $1 - P(E)$, the new node will add another connection randomly.

After nodes have been added or deleted, an agent is sampled to be on turn. We denote this agent by A_t. A_t will add connections excatly as if it was a new agent. After establishing new connections, A_t now updates her opinion based on the opinions of her neighbors.

Let O_1 be the opinion of A_t and O_2 be the opinion of one of A_t's neighbors. Let B be the boundary threshold. The boundary threshold defines when two opinions either pull each other closer together or push each other further apart. For cases where $B \geq |O_1 - O_2|$, we have a case of positive learning where agents pull each other closer. We define a positive learning rate, $PLR \in [0, 1]$. Let V be a fraction of the distance between O_1 and O_2 so that

$$V = (|O_1 - O_2|) \cdot PLR$$

The opinions of the agents are then updated using V. Let O_{max} be the maximum of the two opinions and O_{min} be the minimum of the two opinions:

$$O_{max} = \max(O_1, O_2)$$
$$O_{min} = \min(O_1, O_2)$$

We then pull the two values closer together by V:

$$O'_{max} = O_{max} - V$$
$$O'_{min} = O_{min} + V$$

For cases where $B \leq |O_1 - O_2|$, we have a case of negative learning where agents push each other further away. Similar to the positive learning, we define a negative learning rate $NLR \in [0, 1]$ and let V be given by

$$V = (|O_1 - O_2|) \cdot NLR$$

We then push the values further apart by V:

$$O'_{max} = O_{max} + V$$
$$O'_{min} = O_{min} - V$$

Notice that the only difference between positive and negative learning is the learning rate and whether V pushes or pulls the opinions apart.

The process of updating values is done iteratively for each neighbor of A_t. Notice that updating is symmetric in the sense that both A_t and the neighbor of A_t will be pushed or pulled by V.

After the process of updating is finished for A_t, all ties to agents which are outside of the boundary threshold ($B \geq |O_1 - O_2|$), are disconnected. This concludes one timestep, t, in the model.

B. Model fitting. explain how the model was calibrated (Bayesian Hyperparameter Optimization)

3. Model investigation

Get familiar with the different parameters and their interpretations

A. The effect of randomness. How randomness affects the distribution of opinions, as well as the network

B. The effect of the boundary threshold. How the boundary threshold affects the distribution of opinions, as well as the network

C. The effect of homophily. How homophily affects the distribution of opinions, as well as the network

D. Important interactions. Point to some of the important interactions (possible Golden zones)

4. Results

A. Network generation.

B. Opinion generation.

5. Discussion

6. Conclusion

Use section and subsection commands to organize your document. LaTeX handles all the formatting and numbering automatically.

7. Figures and Tables

Figures and Tables should be labeled and referenced in the standard way using the `\label{}` and `\ref{}` commands.

Figure **??** shows an example of how to insert a column-wide figure. To insert a figure wider than one column, please use the `\begin{figure*}...\end{figure*}` environment. Figures wider than one column should be sized to 11.4 cm or 17.8 cm wide. Use `\begin{SCfigure*}...\end{SCfigure*}` for a wide figure with side captions.

Tables. In addition to including your tables within this manuscript file, PNAS requires that each table be uploaded to the submission separately as a "Table" file. Please ensure that each table .tex file contains a preamble, the `\begin{document}` command, and the `\end{document}` command. This is necessary so that the submission system can convert each file to PDF.

Equations. Authors may use 1- or 2-column equations in their article, according to their preference.

To allow an equation to span both columns, use the `\begin{figure*}...\end{figure*}` environment mentioned above for figures. Using only `\begin{figure*}...\end{figure*}` keeps the equation in a two collum format

Table 1. Character Level Combat Outcomes

	Dependent variable:		
	Combat Amount	Combat Variability	Combat Skill
	(1)	(2)	(3)
Man - Male	0.042***	5.659***	0.031***
	(0.002)	(0.056)	(0.0004)
Woman - Female	−0.026***	1.529***	0.011***
	(0.005)	(0.143)	(0.001)
Woman - Male	0.010	0.375	0.005*
	(0.009)	(0.272)	(0.002)
Player Age	−0.077***		−0.003***
	(0.001)		(0.0002)
Mil. Label	0.135***		0.060***
	(0.002)		(0.0004)
Constant		−97.425***	
		(0.046)	
Char. Order FEs	Y	N	Y
Create Date FEs	Y	N	Y
Observations	576,430	576,430	576,430
R^2	0.028	0.018	0.089

*

p<0.05, ** p<0.01, *** p<0.001

This table reports coefficients and standards errors from ordinary least squares regressions. In all models we can reject the null that *Woman - Female* and *Woman - Male* are equivalent with $p < .01$. In models 2 and 3 we can reject the null that the gender gaps within sex are equivalent ((*Woman - Male*) - (*Woman - Female*) = *Man - Male*) with $p < .001$.

$$\begin{aligned}(x+y)^3 &= (x+y)(x+y)^2 \\ &= (x+y)(x^2+2xy+y^2) \\ &= x^3+3x^2y+3xy^3+x^3.\end{aligned} \quad [1]$$

References

References should be cited in alphabethical order; this will be done automatically via bibtex, e.g. **?**, and **?**. All references should be included in the main manuscript file.

ACKNOWLEDGMENTS. Please include your acknowledgments here, set in a single paragraph. Please do not include any acknowledgments in the Supporting Information, or anywhere else in the manuscript.