Meeting friends of friends and homophily: a complementarity

Adrien Vigier

Received: 18 June 2013 / Accepted: 16 December 2013 / Published online: 25 December 2013
© SAET 2013

Abstract I explore the effects of homophily on the formation of social networks. When individuals are homophilous, friends of friends are likely to share tastes and hence also likely to form new friendships. In the context of homophily, the social network dynamics of meeting friends of friends thus acts as directed search, and a greater number of meetings result in links being formed. However, since it exacerbates preferential attachment—whereby high degree nodes attract new links—homophily also causes more unequal distributions of links. Thus while homophily normally improves social welfare, for a given average number of links formed it in fact affects welfare negatively.

Keywords Social networks · Network formation · Homophily

JEL Classification D85 · A14 · C71 · C72

1 Introduction

Social networks underlie the diffusion of information and form a key component of trust and social capital. In this respect, understanding the structure and determinants of social networks is a major challenge for economists. The present paper aims to shed some light on these topics by exploring some effects of homophily (one's preference for similarity to self) on social networks.

I build in this paper on [Vazquez \(2003\)](#) and [Jackson and Rogers \(2007\)](#) insights on network formation, with a view to illustrate what appears like a natural form of complementarity between homophily and the dynamics of social network formation.

A. Vigier (✉)
Faculty of Economics, University of Oslo, Oslo, Norway
e-mail: adrien.vigier@econ.uio.no

I first postulate, following Jackson–Rogers–Vazquez, that social networks result from a combination of random and network-based (so-called *friends of friends*) meetings. Agents first meet a number of people at random, forming a link to each of them with some probability. They then go on to meet the friends of those they form a link with. If individuals are homophilous then friends of friends are likely to share tastes and hence also likely to form new friendships. Thus, in the context of homophily, the social network dynamics of meeting friends of friends acts as *directed search*. It is in this sense that homophily and the social network dynamics of meeting friends of friends naturally complement one another.

I consider two distinct models of link creation: a homophilous model and a non-homophilous model. The homophilous model is based upon the idea that, in deciding whether or not to form ties with others, individuals consider certain characteristics, and that the same characteristics are used across pairs of individuals. This may be political opinions, taste for certain activities, or simply geographical proximity of two individuals. For our purposes, the important qualification is the following: proximity of tastes is transitive across pairs of nodes. By contrast the non-homophilous model is characterized by the property that the process of link formation is independent across pairs of nodes.

I then go on to show that, in the context of meeting friends of friends, homophily generates networks which tend to (a) have more connections, (b) be more clustered, and (c) exhibit more unequal degree distributions. More connections (and greater clustering) arise because network-based meetings more often result in link formation. This exacerbates *preferential attachment*—whereby high degree nodes attract new links—and explains why homophily also causes more unequal distributions of links. These results have sharp welfare implications. By inducing more links, homophily generates higher average social welfare. However, if agents’ utility is concave in their number of links then, for a given average number of links formed, homophily in fact affects welfare negatively.

Recent years have witnessed significant interest from economists in issues related to the interplay of social network structure and homophily. Bramoulle and Rogers¹ (2010) for instance present a set of complementary views and results to those offered in this paper. They show among other things that, somewhat surprisingly, in multigroup settings network-based meetings tend to have a negative effect on the fraction of links eventually formed within one’s own group. The works of Currarini et al. (2009) and Golub and Jackson (2012) are also to a large extent related to this paper, albeit the focus of the first is on the relationship between group size and homophily rather than network structure *stricto sensu*, while the second investigates the effect of homophily on learning and diffusion in networks. To the best of my knowledge, this paper is the first to bring out and examine the complementarity between homophily and the social network dynamics of meeting friends of friends.

The present paper is organized as follows. I present the formal model in Sect. 2. I develop the analysis and state the paper’s main results in Sect. 3. All proofs are relegated to the Appendix.

¹ “Diversity and Popularity in Social Networks”, unpublished manuscript.

2 Model

Let N denote a set of nodes. Nodes, together with a set of links between pairs of them, define a network g on N . A directed link from node i to node j is represented by g_{ij} : $g_{ij} = 1$ if there is a link going from i to j , and $g_{ij} = 0$ otherwise. If $g_{ij} = 1$ say that j is a neighbor, or friend, of i in network g . The number of nodes to which i is a neighbor is called the *degree*² of node i in network g and denoted by $d_i(g)$, so that $d_i(g) = \#\{j : g_{ji} = 1\}$. Note that it represents the number of nodes from which i can be reached by following a directed link in g . Any network g induces a probability distribution—called degree distribution of g —corresponding to the proportion of nodes having given degree in network g .

I next describe a simple two-step process of network formation: how nodes meet and how, upon meeting, nodes eventually create new links.

Nodes' meetings: The network is formed sequentially. At time $t \in \{0, 1, 2, \dots\}$, node t is added to the population and proceeds to encounter a number of older nodes. He first *meets* m_r other nodes chosen uniformly at random from the existing set of nodes.³ If at this stage t links to t' so that $g_{tt'} = 1$ and t'' is a neighbor of t' then, with probability λ_0 , t also proceeds to meet t'' . The process then repeats itself,⁴ so that if at this stage t links to t'' of whom t''' is a neighbor then, with probability λ_1 , t also proceeds to meet t''' and so on.⁵

Link creation: The paper's main objective is to examine the interplay between social network dynamics (the way meetings occur) and homophily. To this end, I consider two separate models of link creation: a homophilous model, and a non-homophilous model from which to draw comparisons with.

Let u_{ij} denote the net marginal utility gain of i from forming a link with j .⁶ These net utility gains are random variables. I naturally suppose that links are only formed if they generate positive net utility.⁷ Let ϵ^H and ϵ^{NH} denote the probabilities that two randomly chosen nodes form a link upon meeting one another, in the homophilous and non-homophilous models, respectively. Thus $\epsilon^\dagger = \mathbb{P}(u_{ij} \geq 0 | \dagger)$, where $\dagger \in \{H, NH\}$. Both parameters are assumed to lie in $(0, \frac{1}{2})$. I will refer to $\epsilon^\dagger$ as the *base* probability of link formation. Let finally γ denote the circle of circumference 1,

² I thus identify in this paper *degree* and *in-degree*. Indeed, as indicated in the next section, all nodes are treated identically with respect to the number of links they originate (their *out-degree*). So any interesting properties regarding the distribution of links in the network must be related to the in-degree of nodes. More generally, using directed links enhances the tractability of the model and greatly simplifies the analysis. Naturally, once the network formed, directionality can altogether be abandoned by considering the network $\bar{g}$ such that $\bar{g}_{ij} = \max\{g_{ij}, g_{ji}\}$.

³ A meeting may or may not result in a link being formed: the details of this process are described below.

⁴ This process always ends after a finite number of steps since $t''' < t'' < t' < t$.

⁵ We of course restrict attention throughout to sequences $(\lambda_s)_{s \in \mathbb{N}}$ such that the mean-field processes in the models considered converge to a steady-state as $t \rightarrow \infty$. Lemma 2 shows for instance that if $\lambda_s = 0$ for all $s > 0$ then a sufficient condition for the mean-field process to converge is $\frac{3\lambda_0 m_r \epsilon}{4} < 1$.

⁶ In general of course this marginal utility may depend on a number of factors (such as the number and identity of other links that i (or j) has, and so on). I do not pursue this route here.

⁷ I assume here for simplicity that agents are myopic in the sense that they do not account for possible future benefits from creating a link in the form of new meetings.

$\{X_i\}_{i \in N}$ a family of random variables uniformly and independently distributed on γ , and $|\cdot|$ the length of the shortest arc between two points in γ .

Definition 1 (*The non-homophilous model*) Say that nodes are non-homophilous if $\{u_{ij}\}_{i,j \in N}$ constitutes a set of independent random variables.

Definition 2 (*The homophilous model*)⁸ Say that nodes are homophilous if:

$$u_{ij} \geq 0 \Leftrightarrow |X_i - X_j| \leq \frac{\epsilon^H}{2} \quad (1)$$

The homophilous model is based upon the idea that, in deciding whether or not to form ties with others, individuals consider certain characteristics—here represented by X_i —and that the same characteristics are used across pairs of individuals. This may represent political opinions, taste for certain activities, or simply geographical proximity of two individuals.⁹ For our purposes the important qualification is the following: proximity of tastes is transitive across pairs of nodes.

Finally, I follow the literature in assuming that nodes' utility is increasing and concave in the number of links formed and define social welfare as the average utility of nodes in a network.

3 Analysis and results

Given the model's complexity, I follow common practice and base my analysis on mean-field approximation techniques, i.e. I examine the continuous time system where all action happens deterministically at a rate proportional to the expected change.¹⁰

Observe that the probabilities that random meetings result in link creation are given by ϵ^H in the homophilous model and ϵ^{NH} in the non-homophilous model. Let in connection to this, p^H and p^{NH} denote the probabilities that network-based meetings result in link creation. Clearly $p^{NH} = \epsilon^{NH}$. Simple calculations¹¹ establish that $p^H = 3/4 > \epsilon^H$.

Suppose here for simplicity that the process of meeting friends of friends has one round only¹² so that $\lambda_s = 0, \forall s \neq 0$. Let m^H and m^{NH} denote the steady-state average number of links formed by entering nodes in the homophilous and non-homophilous models, respectively. Each entering node first meets m_r other nodes at random and with each of them forms a link with probability $\epsilon^\dagger$. Each of these nodes has on average m friends, which the entering node meets with probability λ_0 . A link is formed with these nodes with probability $p^\dagger$. Thus

⁸ While I borrow the social network dynamics from Jackson and Rogers (2007), the model of homophily used here is inspired by Strogatz and Watts (1998).

⁹ The model trivially extends to allow for multi-dimensional tastes.

¹⁰ The interested reader is referred for example to Vega Redondo (2007) for an extensive discussion of the mean-field approach and to Vazquez (2003) or Jackson and Rogers (2007) for accounts of the performance of mean-field approximation techniques in the specific context investigated in this paper.

¹¹ See Lemma 1 in the Appendix.

¹² Adding more rounds complicates the algebra and provides no further insights.

$$m^{NH} = m_r \epsilon^{NH} (1 + \lambda_0 m^{NH} \epsilon^{NH}) \quad (2)$$

and

$$m^H = m_r \epsilon^H \left(1 + \lambda_0 m^H \frac{3}{4} \right) \quad (3)$$

To allow for meaningful comparisons of the two models we will either fix the base probability of link formation,¹³ i.e. set $\epsilon^H = \epsilon^{NH}$, or fix the average number of links formed,¹⁴ i.e. set $m^H = m^{NH}$. As they provide complementarity insights, I shall make use of both alternatives wherever appropriate.

Let F^H and F^{NH} denote the cumulative distribution functions of the (mean-field) degree distributions resulting, respectively from the homophilous and non-homophilous models. Given cumulative distribution functions F and F' let $F \succ_{FOSD} F'$ denote first-order stochastic dominance of F over F' .

Our first proposition shows that the degree distribution obtained in the homophilous model first order stochastically dominates that obtained in the non-homophilous model. It therefore shows, in the strong sense, that—in the context of meeting friends of friends—for a given base probability of link formation homophily generates more connected networks.

Proposition 1 *Assume that the base probability of link formation is the same in the homophilous and non-homophilous models, i.e. $\epsilon^H = \epsilon^{NH} = \epsilon$. Then $F^H \succ_{FOSD} F^{NH}$. If moreover $\lambda_s = 0$ for all $s > 0$ then for $\dagger \in \{H, NH\}$:*

$$F^\dagger(d) = 1 - \left(\frac{1}{1 + d\lambda_0 p^\dagger} \right)^{\frac{1}{\lambda_0 p^\dagger m_r \epsilon}} \quad (4)$$

I next examine the clustering arising in the two models. Following Jackson and Rogers (2007), I use¹⁵ $C = \mathbb{P}(g_{ij}g_{jk}g_{ik} = 1 | g_{ij}g_{jk} = 1)$. The measure C therefore denotes the fraction of triplets formed per adjacent links in the underlying stochastic process. Let C^H and C^{NH} denote the former measure in the homophilous and non-homophilous models, respectively. Our second proposition confirms the intuitive idea that homophily favors the formation of clusters in networks.

Proposition 2 *If $\lambda_s = 1$ for all $s \in \mathbb{N}$ then $C^H = 3/4 > C^{NH}$.*

Our first results have exploited the idea that, by directing nodes to those which they are most likely to form a link with, network-based meetings are particularly conducive to the formation of new links in the context of homophily. Albeit desirable, this property evidently comes at a cost. By giving more prominence to network-based links, homophily also exacerbates preferential attachment—by which high degree nodes tend to attract a more than proportional share of new links. Thus, for a given average number of links, homophilous networks tend to exhibit more unequal degree

¹³ Note as a preliminary observation that $\epsilon^H = \epsilon^{NH}$ implies $m^H > m^{NH}$.

¹⁴ This implies setting $\epsilon^H = \frac{\epsilon^{NH}}{1 + \lambda_0 m_r \epsilon^{NH} (3/4 - \epsilon^{NH})} < \epsilon^{NH}$.

¹⁵ Or correspondingly for a given network g , $C(g) = \frac{\sum_{i,j,k} g_{ij}g_{jk}g_{ik}}{\sum_{i,j,k} g_{ij}g_{jk}}$.

distributions. Let $F \succ_{SOSD} F'$ denote second-order stochastic dominance of F over F' . We then obtain:

Proposition 3 *Assume that the average number of links formed is the same in the homophilous and non-homophilous models, i.e. $m^H = m^{NH}$. Then $F^{NH} \succ_{SOSD} F^H$.*

I conclude the paper with a record of its welfare implications.

Corollary 1 *For a given base probability of link formation, homophily induces higher social welfare. However, for a given average number of links formed, homophilous networks exhibit lower social welfare.*

Acknowledgments This paper has benefited tremendously from suggestions and comments of people to whom I am greatly indebted: Adam Clark-Joseph, Sanjeev Goyal, Ben Golub, Matthew Jackson, Willemien Kets, Andrea Prat, and Flavio Toxvaerd. In addition, the comments of an anonymous referee have substantially improved the present paper.

Appendix

Lemma 1 *Let X_i, X_j, X_k three random variables uniformly and independently distributed in γ . Then $\mathbb{P}(|X_i - X_k| \leq \frac{\epsilon}{2} | |X_i - X_j| \leq \frac{\epsilon}{2}, |X_j - X_k| \leq \frac{\epsilon}{2}) = \frac{3}{4}$.*

Proof Note that

$$\begin{aligned} & \mathbb{P}\left(|X_i - X_k| \leq \frac{\epsilon}{2} \mid |X_i - X_j| \leq \frac{\epsilon}{2}, |X_j - X_k| \leq \frac{\epsilon}{2}\right) \\ &= \mathbb{P}\left(|X_i - X_k| \leq \frac{\epsilon}{2} \mid |X_i - X_k| \leq \epsilon\right) \end{aligned} \quad (5)$$

while

$$\mathbb{P}\left(|X_i - X_k| \leq \frac{\epsilon}{2} \mid |X_i - X_k| \leq \epsilon\right) = 1 - 2 \int_0^{1/2} \left(1 - \left(\frac{1}{2} + x\right)\right) dx = \frac{3}{4} \quad (6)$$

□

Lemma 2 *Let $\lambda_s = 0, \forall s \neq 0$, and $\frac{3\lambda_0 m_r \epsilon}{4} < 1$. The mean-field dynamics converges to a steady-state both in the homophilous and non-homophilous models.*

Proof We treat both models simultaneously by letting p denote the probability that network-based meetings result in link creation.

Let m_t denote the number of links formed by node t in the mean-field dynamics. We first show that $(m_t)_{t \in \mathbb{N}}$ is a non-decreasing sequence. The proof is by induction. Notice first that $m_0 \leq m_1$, since $m_0 = 0$. Next, suppose the sequence non-decreasing up to $t \geq 1$. Observe that

$$m_t = \varepsilon \left(m_r + m_r \lambda_0 p \frac{1}{t} \sum_{s < t} m_s \right) \quad (7)$$

Hence

$$m_{t+1} - m_t = \varepsilon \lambda_0 p m_r \left(\frac{1}{t+1} \sum_{s < t+1} m_s - \frac{1}{t} \sum_{s < t} m_s \right) \quad (8)$$

By the induction hypothesis, we obtain $m_{t+1} - m_t \geq 0$. So $(m_t)_{t \in \mathbb{N}}$ is a non-decreasing sequence.

Now, by (7), $m_t \leq \varepsilon(m_r + \lambda_0 p m_r \frac{1}{t} t m_t)$ and, upon rearrangement, $m_t \leq \frac{m_r \varepsilon}{1 - m_r \varepsilon \lambda_0 p}$. So $(m_t)_{t \in \mathbb{N}}$ is a non-decreasing and bounded sequence. It therefore converges. Moreover, by (7):

$$\lim_{t \rightarrow \infty} m_t = \frac{m_r \varepsilon}{1 - m_r \varepsilon \lambda_0 p} \quad (9)$$

□

Proof of Proposition 1 For the first part, we show that $F^H(d) = \mathbb{P}^H(d_i < d) < \mathbb{P}^{NH}(d_i < d) = F^{NH}(d)$, for all d . For $\epsilon^H = \epsilon^{NH}$ the number of links formed at random is the same in the two models, while we can deduce from Lemma 1 that the number of links formed through network-based meeting is strictly greater under homophily. The result is thus readily obtained by conditioning on the number of links formed at random.

For the second part, as in the proof of the previous lemma I treat both models simultaneously by letting p denote the probability that network-based meetings result in link creation.

Let m denote average degree in the steady-state of the mean-field dynamics, and $d_i(t)$ the in-degree of node i at time t . The probability that t meets i at random is $\frac{m_r}{t}$. Upon meeting in this way the probability of a link being formed from t to i is ϵ . In addition, t may meet i through the network. The total number of nodes reached in this way is $m_r \epsilon \lambda_0 m$, and i is among these nodes with probability $\frac{d_i(t)}{mt}$. Meeting in the way, the probability of a link being formed from t to i is p . Summing up, the total probability of a link being formed from t to i is given by $\frac{m_r \epsilon}{t} + m_r \epsilon \lambda_0 m \frac{d_i(t)}{mt} p$. We thus obtain

$$\frac{dd_i(t)}{dt} = \frac{m_r \epsilon}{t} + \left(\frac{\lambda_0 m_r \epsilon p}{t} \right) d_i(t) \quad (10)$$

Which, using the boundary condition $d_i(i) = 0$, gives

$$d_i(t) = \frac{1}{\lambda_0 p} \left(\left(\frac{t}{i} \right)^{\lambda_0 m_r \epsilon} - 1 \right), \quad t \geq i \quad (11)$$

Next, let $i(d, t)$ denote the unique node with degree d at time t . By (11) we have $i(d, t) = t \left(\frac{1}{1 + d \lambda_0 p} \right)^{\frac{1}{\lambda_0 p m_r \epsilon}}$. If $F_t(\cdot)$ denotes the cumulative distribution function of the in-degree distribution at time t we then have $1 - F_t(d) = \frac{i(d, t)}{t}$. Simple rearrangement now yields (4). □

Proof of Proposition 2 Note first that given $\lambda_s = 1, \forall s \in \mathbb{N}$, any node has met all friends of his friends. Hence, by definition of the homophilous model

$$\begin{aligned} & \mathbb{P}(g_{ij}g_{jk}g_{ik}=1|g_{ij}g_{jk}=1) \\ &= \mathbb{P}\left(|X_i - X_k| \leq \frac{\epsilon}{2} \mid |X_i - X_j| \leq \frac{\epsilon}{2}, |X_j - X_k| \leq \frac{\epsilon}{2}\right) \end{aligned} \quad (12)$$

with X_i, X_j, X_k uniformly and independently distributed on γ . Using Lemma 1 thus gives $C^H = \frac{3}{4}$.

By contrast links are totally uninformative in the non-homophilous model, whence $\mathbb{P}(g_{ij}g_{jk}g_{ik}=1|g_{ij}g_{jk}=1) = \epsilon$. $\square$

Proof of Proposition 3 The proposition follows immediately from Proposition 1 here in conjunction with theorem 6 of Jackson and Rogers (2007) and the observation that $m^H = m^{NH}$ implies $r^H < r^{NH}$, where r denotes the ratio of the number of links formed at random to the number of links formed through the network. $\square$

References

- Curranini, S., Jackson, M., Pin, P.: An economic model of friendship: homophily, minorities and segregation. *Econometrica* **77**(4), 1003–1045 (2009)
- Golub, B., Jackson, M.O.: How homophily affects the speed of learning and best-response dynamics. *Q. J. Econ.* **127**(3), 1287–1338 (2012)
- Jackson, M., Rogers, B.: Meeting strangers and friends of friends: how random are social networks. *Am. Econ. Rev.* **97**(3), 890–915 (2007)
- Strogatz, S., Watts, D.: Collective dynamics of ‘Small-World’ networks. *Nature* **393**, 440–442 (1998)
- Vega-Redondo, F.: *Complex social networks*. Cambridge University Press, Cambridge (2007)
- Vazquez, A.: Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. *Phys. Rev. E* **67**(5), 056104 (2003)