

Empirical Models of Social Learning in a Large, Evolving Network

Ayse Bener, Ryerson University
Bora Çağlayan, Ryerson University
Adam Douglas Henry, University of Arizona
Marek Lipczak, Google Corporation
Pawel Pralat, Ryerson University

Manuscript revised November 2015

Abstract

This paper advances theories of social learning through an empirical examination of how social networks—which constrain individuals’ access to information about the behaviors and cognitions of other people—evolve over time. Using data on a large social network of mobile device users over a one-month time period, we test three hypotheses: 1) attraction homophily causes individuals to form ties on the basis of attribute similarity, 2) aversion homophily causes individuals to delete existing ties on the basis of attribute dissimilarity, and 3) social influence processes cause individuals to adopt the attributes of others they share direct ties with. Statistical models offer varied degrees of support for all three hypotheses and show that these mechanisms are more complex than assumed in prior work. Although homophily is normally thought of as a process of attraction, people also avoid relationships with others who are different. These mechanisms have distinct effects on network structure. And while social influence does help explain behavior, people tend to follow global trends more than they follow their friends.

Introduction

What determines the adoption of beliefs, new technologies, cultural behaviors, and cooperation among human agents? These are all questions of *social learning*, a term that refers broadly to the processes by which a person's social environment shapes their actions (how a person behaves) and cognitions (how a person thinks). Through social learning, an agent—a term we use to refer to any individual actor within a social network—may adopt the cognitions or behaviors from those they have an opportunity to observe or interact with directly.

Combined with forms of individual learning, where an agent learns through their own direct experiences, social learning may provide a more efficient, less costly, and less risky way to learn optimal social behaviors (Feldman, Aoki, & Kumm, 1996; Olsson & Phelps, 2007; Rendell et al., 2010; Sandberg et al., 2012). Developing a better understanding of the mechanisms of social learning can yield insights into the development and evolution of human culture. At the same time, theories and models of social learning are also of great practical importance as they can also inform how people can more effectively collaborate to solve wrenching social, environmental, and economic problems, such as global climate change (Adger, 2005; Bidwell, Dietz, & Scavia, 2013; Gerlak & Heikkila, 2011; Henry, 2009; Parson & Clark, 2005).

Social networks will occupy a prominent role in a comprehensive theory of learning. This is because direct social contacts—who people interact with, talk to, and observe on a regular basis—provide the raw material needed for learning and define a large part of the social environment that constrains what a person could *potentially* learn from other social agents (Henry 2009; Friedkin & Johnsen, 1997; Friedkin, 2001; Marsden & Friedkin, 1993). At the same time, products of learning (behaviors and cognitions) also shape the social environments of

human agents, since individuals also exercise control over their social environment and potentially select network partners as a function of individual attributes, including their behaviors and cognitions (Boardman, Domingue, & Fletcher, 2012; Girard, Hett, & Schunk, 2015; Jackson, 2008; McPherson, Smith-Lovin, & Cook, 2001). In this way, social network structures co-evolve with the products of social learning (Aral, Muchnik, & Sundararajan, 2009; Dandekar, Goel, & Lee, 2013; Lewis, Gonzalez, & Kaufman, 2012). A better understanding of social learning processes therefore depend on understanding how social agents form networks, are influenced by others within their network, and how these two phenomena work in tandem.

This paper contributes to our understanding of social learning by developing and testing empirical models of how social networks evolve over time; particularly, how people in a social network choose links on the basis of their own attributes (a phenomenon known as homophily), and how individual attributes are in turn shaped by network structure (a phenomenon known as social influence). Examining these co-evolutionary processes is challenging because it requires data on both networks and attributes that shift over time. So while there are many theoretical models of homophily and social influence, empirical datasets that allow scholars to test these models and calibrate theoretical expectations based on empirical findings are relatively rare. There are, however, a number studies that have tested for social influence and homophily within a single social network, such as Kandel's (1978) now-classic study of adolescent friendships, homophily, and socialization in terms of behaviors (drug use and delinquency) and cognitions (political ideology and education goals). Other more recent studies include Lazer's (2001) study of attitudes and communication among federal agency bureaucrats, Aral, Muchnik, & Sundararajan's (2009) study of product adoption and

instant messaging among mobile device users, and Lewis, Gonzalez, & Kaufman's (2012) study of tastes and online social networks among college students.

We build on these studies through an analysis of a large social network of approximately 300,000 mobile device users, for whom data are available on both social contacts as well as individual behaviors—measured as the “tastes” that users exhibit in their use of the Web and mobile applications. In using these data we build on related studies that have used data on tastes and virtual connections (i.e., through online friendships or mobile communications) to examine fundamental social processes occurring in networks (Eagle, Pentland, & Lazer, 2009; Lewis, Gonzalez, & Kaufman, 2012). By observing how contacts (networks) and tastes (attributes) change over a one-month time period, we are able to explicitly test hypotheses of network evolution that cannot be examined by looking at cross-sectional data alone.

Mechanisms of Social Learning: Homophily and Social Influence

We study the coevolution of network structure and nodal attributes by focusing on three particular social processes as depicted in Figure 1. A distinction here is made between those variables that are broadly thought to interact with one another (variables existing at the framework level—yellow boxes in Figure 1) and the particular mechanisms through which the variables cause change in one another (factors existing at the model level—green boxes in Figure 1). At the framework level, social networks and actor attributes are viewed as having a reciprocal influence on one another. The particular processes at the model level include the formation and deletion of ties based on similarity or differences in agent attributes

(homophily), and that certain attributes are likely to diffuse through existing network ties (social influence).

[FIGURE 1 ABOUT HERE]

The role of homophily

Actor attributes influence the structure of networks through homophily, a widely-studied phenomenon where actors tend to position themselves in networks such that they are connected to others similar to themselves (Jackson, 2008; McPherson, Smith-Lovin, & Cook, 2001). That actors form ties on the basis of homophily along a wide spectrum of traits is one of the most robust empirical findings in the social networks literature. In social networks, homophily has been observed to operate on socio-economic variables such as race (Currarini, Jackson, & Pin, 2009, 2010), nationality (Girard, Hett, & Schunk, 2015), and gender (Kleinbaum, Stuart, & Tushman, 2013; Stehlé et al., 2013), cognitive and behavioral traits such as personal tastes (Lewis, Gonzalez, & Kaufman, 2012) and one's propensity to cooperate with others (Girard, Hett, & Schunk, 2015), as well as physical genetic factors (Boardman, Domingue, & Fletcher, 2012). Homophily has also been studied at the level of social aggregates. For example, it is hypothesized that organizations involved in the policy process tend to coordinate with other organizations sharing their values, interests, and beliefs about policy problems (Sabatier & Jenkins-Smith, 1993; Weible et al., 2011)—a hypothesis that is supported by numerous empirical studies (e.g., Henry et al., 2011; Ingold & Fischer, 2013; Leach & Sabatier, 2005; Weible & Sabatier, 2005). Similarly, homophily has been shown to operate at the level of entire cities, such that municipal governments choose other cities as collaborators on the basis of

similarity in socio-economic traits as well as shared political ideology (Gerber, Henry, & Lubell, 2013).

Given the widespread recognition that network structures are determined in part by homophily, then why continue to study this phenomenon? There are at least four answers to this question. First, factors other than homophily help to determine network structure (Jackson, 2008; Snijders et al., 2006), and an analysis of homophily effects must therefore consider how homophily operates in tandem with other structural influences such as optimization (D'Souza et al., 2007), seeking out connections with high-degree nodes (Aiello et al., 2008; Barabási & Albert, 1999), and the tendency to maintain connections within closed triads, or among “friends of friends” (Burt, 2000; Carpenter, Esterling, & Lazer, 2004; Krackhardt, 1999).

Second, the theoretical mechanisms that lead to homophily in certain contexts are not fully understood. One persistent question in understanding homophily is whether this phenomenon is due to a systematic bias for ties with similar individuals (known as “choice homophily”) or due to the increased opportunities for forming ties with similar individuals because people select themselves into social contexts—such as schools, clubs, or residential neighborhoods—based on certain traits (known as “structural homophily”). Clearly distinguishing between these effects, and making clear inferences as to whether observed correlations between network ties and attribute similarity is due to choice or structural homophily, is an important area of research (Boucher, 2015; Bramoullé et al., 2012; Currarini, Jackson, & Pin, 2010; Kleinbaum, Stuart, & Tushman, 2013; Kossinets & Watts, 2009).

Third, insofar as choice homophily is concerned, the reasons why people tend to select network partners with similar traits are not well understood. Kossinets & Watts (2009: 406)

offer the intuitive explanation that “people form ties with similar others because, rightly or wrongly, they prefer to.” Theoretical explanations for this rely on the idea that the creation and maintenance of network ties is costly (D’Souza et al., 2007; Saramaki et al., 2014), and these “transaction costs” associated with social ties are reduced when one interacts with others who are similar (Feiock, 2013). Having more in common with one’s social partners makes it easier to form and maintain healthy relationships. Of course, this is also complicated by the fact that different traits might matter differently for homophily, and these differences are likely to be highly dependent on context—so, for example, a person might select their friends based on similarity in political views, but not based on similarity in musical tastes.

Choice homophily, as we show here, is not just a matter of tie formation. Choice homophily can also influence choices about the termination of social ties, or avoidance of certain relationships altogether. We refer to these processes respectively as “attraction homophily” and “aversion homophily.” While attraction and aversion homophily might appear to be two sides of the same coin, in fact these are two distinct mechanisms that might operate independently of one another (Noel & Nyhan, 2011). To illustrate, consider Thomas Schelling’s classic work on the emergence of residential segregation (Schelling, 1971). Starting with the observation that many communities exhibit racial segregation (such that people of similar races tend to live spatially proximate to one another) a natural explanation is that people have a strict preference for living in a racially homogenous neighborhood—that is, that individuals exhibit *attraction homophily* in their choice of neighborhood, based on race. On the other hand, Schelling’s model demonstrated that one need not assume that people seek out homogenous neighborhoods for segregation to emerge. Rather, it is sufficient to assume that individuals

have a slight bias against being a minority in their local neighborhood—that is, people exhibit a small amount of *aversion homophily* in their choice of residential neighborhood, based on race. When one models these processes in a network context it can be shown that slight levels of aversion homophily are sufficient to produce global patterns of network segregation and community structure that are characteristic of many real-world networks (Henry, Prałat, & Zhang, 2011).

In this paper, we view both types of homophily as independent processes that influence the structure of observed networks: first through the deletion of ties as a result of aversion homophily (Figure 1, process #1), and second, through the formation of new ties as a result of attraction homophily (Figure 1, process #2). As noted above, we study homophily in terms of the various tastes that people have regarding Web browsing behavior and mobile application usage. We therefore consider two homophily hypotheses:

H₁: In a social network, the likelihood of tie formation between two actors increases with greater similarities in the actors' attributes.

H₂: In a social network, the likelihood of tie deletion between two actors increases with greater differences in the actors' attributes.

The role of social influence

While actor attributes—such as tastes—shape network structures, networks also have a reciprocal effect on attributes. Many attributes that drive network structures are malleable,

particularly cognitive and behavioral attributes such as opinions, values, and cultural traits. Through a process of social influence, the attributes of a particular actor's network neighbors (those nodes that the actor shares direct connections with) potentially shape the focal actor's attributes (Friedkin & Johnsen, 1997; Friedkin & Johnsen, 2011; Friedkin, 2001). Social influence is depicted as process #3 in Figure 1.

Studies of social influence are grounded in the observation that one's social network partly determines one's actions and behaviors. "Social networks" in this sense need not be actual, physical connections, but merely pathways that provide information about other actors' behavior or cognitions (Marsden & Friedkin, 1993). Scholars have long been interested in developing models of social influence, with early applications in the modeling of how groups with varied opinions on an issue eventually reach a consensus (DeGroot, 1974). Studies of social influence are not limited to an investigation of why people hold certain opinions, attitudes, and beliefs (the focus of prominent social influence scholars such as Friedkin & Johnsen, 2011). Models of social influence may be applied to related phenomena such as the adoption of new interests and tastes in an online network (Lewis, Gonzalez, & Kaufman, 2012), new communications technologies (Aral, Muchnik, & Sundararajan, 2009), sustainability behaviors (Bollinger & Gillingham, 2012; Henry & Vollan, 2014), and even the diffusion of innovations across organizations (O'Neill, Pouder, & Buchholtz, 1998) and governments (Berry & Berry, 2007).

In terms of people's tastes, we hypothesize that social influence processes will cause a convergence in tastes among those connected in a social network, giving our third main hypothesis:

H₃: In a social network, actors tend to adopt the attributes of others they share direct connections with.

It should be noted that this hypothesis should not apply to all situations where actors exercise—or are prone to—social influence in a network setting. For instance, when actors are learning about scientific information characterized by uncertainty, then actors may have differential levels of trust in different information sources due to biased assimilation (Innes, 1978; Lord, Ross, & Lepper, 1979; Munro & Ditto, 1997), which may create and reinforce polarization in a social network (Dandekar, Goel, & Lee, 2013).

Another important caveat to this hypothesis is that social influence effects are likely to be mediated by various properties of the network. For instance, social influence (as the degree of convergence in tastes between two actors) is likely to be stronger when actors have higher levels of trust, or more frequent interactions. Actors are therefore more likely to adopt the tastes of others who are more similar to themselves, or who one shares both direct and indirect connections with (Burt, 2000; Centola, 2011). Moreover, actors are not only influenced by those they share direct connections with, but are likely also influenced by global shifts as well. So while those in a local network neighborhood likely exert the strongest influences on tastes, shifts in overall tastes—e.g., changes in what is “trending” or “in fashion”—are also likely to be a source of social influence. This possibility is examined in the analysis of a large empirical dataset on social connections and tastes, examined in the next section.

Data Source

To test the above hypotheses of homophily and social influence in networks, we examine a unique dataset on instant messaging and user tastes from a popular mobile phone platform in the United States. In this analysis, network ties represent contacts between individual mobile phone users—more specifically, if user A has user B in their phone’s instant messaging contact list, then we form an undirected, unweighted network tie from A to B.

Attributes of network nodes are viewed as a combination of Web browser activity and application download history. These behaviors are summarized by a multi-dimensional attribute called “tastes,” which represents the overall interests and behavior of users—at least as they relate to how individuals use their mobile devices. The various dimensions of the taste attribute, called taste categories, were developed by the mobile platform developers and align with the platform’s application categories. While taste categories form a tree structure, in this paper we examine only the seventeen top-level taste categories: lifestyle, finance, business, entertainment, ringtones, photo, news, utilities, health, sports, themes, books, games, social, productivity, education and navigation. A proprietary algorithm is used to assign taste scores between 0 and 1 for individual users over time in each of these taste categories, where higher values indicate stronger interests in Web searches and applications relevant to the given category. This provides a dynamic view of how the interests and behaviors of users change over time.¹

¹ It should be noted that tastes are only estimated for users with activity in a given month, and for those users who allow their usage history to be tracked (privacy settings allow users to prevent their usage history from being

The dataset used here is drawn from two snapshots of the contact network and user tastes (attributes) at two points in time: October 19, 2013 and November 22, 2013. A one-month time interval was chosen so that we might observe short-term changes in the network—giving us a view of what might be thought of as the “first derivative” of the longer-term changes in attributes and social networks over time.

Included in this analysis are all mobile platform users in the United States with taste data in a given month—this excludes all users who either had no recorded activity, as well as those users who used their privacy settings to prevent tracking of their usage history. This yields network and attribute data on approximately 300,000 active users, which represents about 10% of all users in the United States between these time periods.

Table 1 provides descriptive statistics for the two network snapshots, and degree distributions for both networks are depicted in Figure 2. Between the two time periods, there were approximately 426,000 removals of ties and 1.1 million new ties among the active users studied here. If we examine the full network (and not only the users included in this analysis), we see that the network grew in size between the two time periods. Despite this 12% increase in the number of nodes in the network, however, the basic network characteristics remained stable. As depicted in Figure 2, both network snapshots are heavy-tailed (with a maximal degree of about six thousand), sparse networks (with an average degree of 3.15), with a large giant connected component (containing 85% of all nodes), and a large number of smaller disconnected components. These characteristics are typical for communication networks among mobile phone users.

recorded). have not set their privacy settings opted to send usage history and in addition some users do not have recorded taste opted to set privacy settings some have no taste information recorder users

[TABLE 1 ABOUT HERE]

[FIGURE 2 ABOUT HERE]

Results for Homophily

The structure of these data allow us to explicitly test two distinct views of homophily discussed above—homophily as a process of *attraction* to others with *similar* attributes (H_1), and homophily as a process of *aversion* from others with *dissimilar* attributes (H_2). To do this, we take all dyads (i.e., pairs of individual actors) in the network and group them into one of four categories as summarized in Table 2. These groups define whether the dyad contained a newly formed tie (Group 0-1), a deleted tie (Group 1-0), a tie that was present in both time periods (stable ties; Group 1-1), or a tie that was absent in both time periods (stable non-ties; Group 0-0).

[TABLE 2 ABOUT HERE]

For each of the dyads in the system, we also calculate two alternative measures of how different each pair of actors are in terms of their attributes. In particular, for network nodes A and B, the *taste difference* between A and B is defined as either the Euclidean distance or the Hamming distance between A and B's respective tastes (as noted above, each individual taste is a seventeen-dimensional vector).² While Hamming distance is estimated based on the number of shared tastes, Euclidean distance is estimated by weighting the shared tastes. Since tastes do

² We use a measure of difference rather than similarity because zero has a natural interpretation in a difference measure (i.e., zero means that two nodes have exactly the same attribute).

change between the two time periods, the *taste difference* variable is recalculated for each snapshot.

Given both the taste difference measures and the grouping of dyads, we are able to test the homophily hypotheses according to the following logic:

- 1) Under the attraction hypothesis (H_1): Among all pairs that are *not connected* in the first time period (Groups 0-0 and 0-1), those pairs with greater similarity in tastes are more likely to form a tie between the two periods (that is, belong to Group 0-1).
- 2) Under the aversion hypothesis (H_2): Among all pairs that are *connected* in the in the first time period (Groups 1-1 and 1-0), those pairs with greater dissimilarity in tastes are more likely to delete a tie between the two periods (that is, belong to Group 1-1).

Homophily as attraction

The attraction hypothesis is tested by fitting a logistic regression model using dyads in Groups 0-0 and 0-1 as the unit of analysis:

$$Y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{taste_diff} + \beta_2 \text{path_length})}}, \text{ where} \quad (1)$$

- Y is the probability of tie formation (that is, whether the dyad is a member of Group 0-0 versus Group 0-1),
- β_0 , β_1 , and β_2 are constant coefficients,
- Variable *taste_diff* is the taste difference between the two nodes in the dyad in the first time period, measured either as Euclidean distance or Hamming distance between taste vectors as described above, and

- Variable *path_length* is the geodesic path length between the two nodes in the dyad in the first time period (when all dyads included in this model are empty). This variable is included as a control, to capture the possible effects of being more proximate in the social space on tie formation. For example, a large body of research on triadic closure suggests that network actors are more likely to form ties if they both have ties to a common third actor (in which case the value of *path_length* would be two).

Two important issues with this approach require discussion. First, the use of logistic regression models assume that observations are independent of one another—in other words, the probability of tie formation between one pair of actors is uncorrelated with tie formation between any other two actors in the system. This is not a good assumption, especially given that the same actors appear in multiple dyads. To deal with this problem, we fit models only on a random sample of dyads in the network. Given that this is a relatively small percentage of actor-actor pairs spread out over an entire national market, the chances that we have replicated agents in the sample is very small, and we will be able to safely assume that dyads included in this analysis are, for the most part, independent.

The second issue is that Group 0-0 is extremely large compared to Group 0-1. By comparing samples drawn from both groups we invoke an implicit theory that all disconnected actor pairs have an equal opportunity to form a link between the two time periods. This is, of course, unrealistic. Thus, we are in need of a way to narrow the population of stable non-tie dyads (Group 0-0) to a subgroup of dyads that we can realistically assume had an opportunity to form a link between the two time steps.

Our approach is to look at the geodesic path distances between actor pairs in Group 0-0 in the first time step, and compare them with the geodesic path distances between actor pairs in Group 0-1 (newly formed ties). These distributions, depicted in Figure 3, suggest that there is a maximum path length beyond which it can be safely assumed that actors are in completely different communities, with no opportunity to form ties with one another.

[FIGURE 3 ABOUT HERE]

Examining differences between these distributions enables us to develop a heuristic, based on path length, for whether two actors have an opportunity to form a tie. The 98th percentile of path lengths among actors with newly-formed ties (Group 0-1) is six, meaning that fewer than two percent of all actors who formed a tie started with greater than six degrees of separation. This suggests that dyads excluding from Group 0-0 all dyads where actors are separated by more than 6 degrees of separation narrows the universe of Group 0-0 dyads to those who plausibly had an opportunity to form a tie.

Table 3 summarizes the results obtained from fitting the logistic regression model (equation 1) after the exclusion of Group 0-0 dyads with path length seven or greater, and a random sampling of 20% of all remaining dyads from each group. As predicted (H_1), we find strong negative coefficients on both concepts of taste differences. This indicates that as taste differences in the first time period increase, the probability of tie formation between the two time periods decreases. This effect is strongly significant ($p < 0.001$ in both models).

[TABLE 3 ABOUT HERE]

Homophily as aversion from those who are different

Testing for the aversion aspect of homophily is more straight-forward, since it is not necessary to estimate the population of ties that are candidates for deletion—dyads in Groups 1-0 and 1-1 all have an existing tie on the first time period, all of which are treated as candidates for deletion. Similar to the analysis reported above, the probability of tie deletion is estimated as a logistic function of attribute differences in the first time period, using dyads drawn from Groups 1-0 and 1-1 as the unit of analysis:

$$Z = \frac{1}{1+e^{-(\beta_0+\beta_1 \text{ taste_diff})}}, \text{ where} \quad (2)$$

- Z is the probability of tie deletion (that is, whether the dyad is a member of Group 1-0 versus Group 1-1),
- β_0 and β_1 are constant coefficients, and
- Variable *taste_diff* is the taste difference as described above.

This logistic regression model (2) is estimated using another random sample of 20% of all dyads to deal with the potential problem of interdependence among observations. Results from this logistic regression analysis are reported as Models 3 and 4 in Table 3.

As with H_1 , these results lend strong support for the aversion hypothesis (H_2). Note that the effect of both taste difference measures—Euclidean distance and Hamming distance—are both positive and significantly different from zero ($p < 0.001$ in both models). This indicates that,

as distances increase, the probability of tie deletion also increases.³ This effect is also represented graphically in Figure 4.

[FIGURE 4 ABOUT HERE]

Results for Social Influence

The longitudinal nature of this dataset also allows us to directly test the social influence hypothesis (H_3), that individuals tend to adopt the attributes of those they are connected to in a social network. In the context of this study, we examine whether changes in tastes between the two time periods can be explained by social contacts in the first time period.

The approach used here is to view shifts in each user's tastes between the two time periods as a function of two factors: 1) shifts in the average tastes of the individual that the user is connected to (which we refer to as the user's *local center of mass*), and 2) global shifts in the average tastes of all users (which refer to as the system's *global center of mass*). More formally, for a single taste element T , a particular user's change in tastes is modeled as follows:

$$T_2 - T_1 = a(L_2(T) - L_1(T)) + b(G_2(T) - G_1(T)) + e, \text{ where} \quad (3)$$

- T_1 and T_2 are the user's tastes during the first and second time periods,
- $L_1(T)$ and $L_2(T)$ are the user's local centers of mass at the first and second time periods for taste element T (that is, the average taste T of individuals the user is connected to in a given time period),

³ The models including Hamming distance allows for a slightly easier interpretation, since coefficients in these models tell us how dissimilarity in each additional taste dimension will change the probability of tie formation or deletion.

- $G_1(T)$ and $G_2(T)$ are the global centers of mass for taste T in the first and second time periods,
- a and b are constant coefficients ranging from 0 to 1, inclusive, and
- e is an error term capturing motivations for change not accounted for in the other terms.

According to this model, we assume that a user's shift in tastes will follow some constant fraction of the shift in the global center of mass (that is, users will "follow" global trends to some degree), and at the same time users will also follow trends within their local network neighborhood. The constant coefficients a and b are weights that indicate the degree to which a user is influenced by these global and local trends, where a zero value of a or b indicates that users are not at all influenced by shifts in local or global tastes, respectively, and a value of one indicates that users are strongly influenced by shifts in these tastes. Finally, the error term, e , captures a user's individual propensities for changing tastes, that is, factors other than one's social environment as measured by the L and G variables.

Since the model above (3) expresses changes in user tastes between the two time periods as a linear combination of observed local and global shifts in tastes, the coefficients a and b may be estimated using a standard ordinary least squares (OLS) regression. Table 4 reports the results of seventeen regression models, one for each of the measured user tastes. While the goodness of fit varies greatly across tastes, the estimated weights on local and global shifts in tastes (a and b , respectively) are significantly different from zero in all models ($p < 0.001$). This suggests that the tastes of users are indeed influenced by their social environment.

[TABLE 4 ABOUT HERE]

As noted above, other aspects of the network may exercise an influence over these processes of social influence; that is, as a function of their local network, some users may be more or less susceptible to social influences. One such influence is the degree of nodes—it turns out that users with more contacts also tend to experience smaller shifts in their own tastes. This pattern is depicted graphically in Figure 5, which shows for a given taste the distribution of users' change in tastes between the two time periods (vertical axis) against the degree of the user in the first time period (horizontal axis). Shifts in four tastes are shown here for illustrative purposes; however, this decreasing trend is seen across all measured tastes. This suggests that as individuals are exposed to greater numbers of social contacts, they tend to be less influenced by their social environment.

[FIGURE 5 ABOUT HERE]

Conclusion

This paper contributes to our understanding of social learning and network evolution through the analysis of a large mobile contact network, coupled with data on the tastes of individuals, over two points in time. The unique structure of these data allow for an examination of how tastes influence the formation of new contacts (attraction homophily, H_1) and the deletion of contacts (aversion homophily, H_2), and how contacts in turn influence the tastes of actors in the network (social influence, H_3).

We find evidence in support of all three hypotheses. In terms of attraction homophily, the overall similarity between actors in terms of their tastes is a significant predictor that they will form a tie, controlling for the degree of separation between these actors in the network.

Similarly, aversion homophily is seen in that among connected actors, those that terminate ties tend to be more different in terms of their tastes than those connected actors who do not cut their ties. Finally, there is strong evidence also that tastes diffuse through existing links—that is, that shifting individual tastes do tend to follow changes in the average tastes of one’s network contacts. Interestingly, shifts in global trends appear to explain far more of the variation in individual tastes than shifts in the tastes of one’s direct network contacts. This finding is contrary to expectation for network scholars who argue that people are more strongly influenced by those they are socially close to. It seems that trends in local neighborhoods do matter—that is, social influence does shape tastes—however these data suggests that people follow trends more than they follow their friends.

This paper builds on a large and vibrant literature on network evolution. This research shows that it is useful for scholars—both in theoretical models as well as in empirical work—to make a distinction between the rationales for tie formation and tie deletion. While in this dataset we find similar strengths of taste similarity on the formation and deletion of ties, in other contexts homophily may not operate in both directions. It may be possible that actors will exhibit a small preference for homophily in the creation of new ties networks, but the maintenance of ties with dissimilar alters may be very costly.

We also show that social influence is a complex phenomenon that is unlikely to reduce to a simple process of averaging across the tastes (or other attributes) of one’s network neighbors. The effect of social influence differs greatly depending on the attribute in question. In the case of tastes, some attributes exhibit relatively low levels of social influence while others are quite prone to change as a result of local and global shifts. Social influence is also

mediated by other factors such as degree; those who have relatively few connections tend to be, on average, much more strongly influenced by their local network neighborhood.

Finally, this research illustrates the potential of working with large datasets on contacts and mobile device usage. The strength of the homophily and influence effects seen in this research suggest that we are, indeed, examining a dataset that lend insights into other types of social networks—for example, networks involving actual physical interactions between people, who vary their worldviews and behaviors based on those they interact with socially.

At the same time, it is important in future work to explicitly examine the degree to which datasets such as these are appropriate laboratories to test hypotheses of social learning and network evolution. This dataset allows for an examination of how the tastes of mobile device users influence the tastes of any particular user. However, this is probably just one part of the true social environment that explains tastes. Similarly, in the future it will be important to examine a more broad and fine-grained set of attributes that both determine and are influenced by network structure. As defined here, tastes are one useful measure of human behavior, but future work should aim to extend this to attribute variables that help us to further understand salient environmental, economic, and political problems—such as understandings of scientific information, political ideology, or norms of cooperative behavior.

Overall, in this research we contribute to theories of social learning by showing how learning is at least partly conditional on social network structures, which both constrain and enable the exposure of individual agents to new information, technologies, and behaviors. As we show in this paper, an essential part of a theory of learning will be robust models of how

network structures co-evolve with learned attributes, with a particular emphasis on the roles of homophily—both aversion and attraction homophily—and social influence.

Literature Cited

- Adger, W. N. (2005). Social-Ecological Resilience to Coastal Disasters. *Science*, 309(5737), 1036–1039.
- Aiello, W., Bonato, A., Cooper, C., Janssen, J., & Prałat, P. (2008). A Spatial Web Graph Model with Local Influence Regions. *Internet Mathematics*, 5(1), 175–193.
- Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks. *Proceedings of the National Academy of Sciences*, 106(51), 21544–21549.
- Barabási, A.-L., & Albert, R. (1999). Emergence of Scaling in Random Networks. *Science*, 286(5439), 509–512.
- Berry, F. S., & Berry, W. D. (2007). Innovation and Diffusion Models in Policy Research. In *Theories of the Policy Process* (2nd ed., pp. 223–260). Boulder, Colorado: Westview Press.
- Bidwell, D., Dietz, T., & Scavia, D. (2013). Fostering Knowledge Networks for Climate Adaptation. *Nature Climate Change*, 3.
- Boardman, J. D., Domingue, B. W., & Fletcher, J. M. (2012). How Social and Genetic Factors Predict Friendship Networks. *Proceedings of the National Academy of Sciences*, 109(43), 17377–17381.
- Bollinger, B., & Gillingham, K. (2012). Peer Effects in the Diffusion of Solar Photovoltaic Panels. *Marketing Science*, 31(6), 900–912.
- Boucher, V. (2015). Structural Homophily. *International Economic Review*, 56(1), 235–264.
- Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P., & Rogers, B. W. (2012). Homophily and Long-Run Integration in Social Networks. *Journal of Economic Theory*, 147(5), 1754–1786.
- Burt, R. S. (2000). The Network Structure of Social Capital. *Research in Organizational Behavior*, 22, 345–423.
- Carpenter, D. P., Esterling, K. M., & Lazer, D. M. J. (2004). Friends, Brokers, and Transitivity: Who Informs Whom in Washington Politics? *Journal of Politics*, 66(1), 224–246.
- Centola, D. (2011). An Experimental Study of Homophily in the Adoption of Health Behavior. *Science*, 334(6060), 1269–1272.
- Currarini, S., Jackson, M. O., & Pin, P. (2009). An Economic Model of Friendship: Homophily, Minorities and Segregation. *Econometrica*, 77(4), 1003–1045.
- Currarini, S., Jackson, M. O., & Pin, P. (2010). Identifying the Roles of Race-Based Choice and Chance in High School Friendship Network Formation. *Proceedings of the National Academy of Sciences*, 107(11), 4857–4861.
- Dandekar, P., Goel, A., & Lee, D. T. (2013). Biased Assimilation, Homophily, and the Dynamics of Polarization. *Proceedings of the National Academy of Sciences*, 110(15), 5791–5796.
- D'Souza, R. M., Borgs, C., Chayes, J. T., Berger, N., & Kleinberg, R. D. (2007). Emergence of Tempered Preferential Attachment from Optimization. *Proceedings of the National Academy of Sciences*, 104(15), 6112–6117.
- Eagle, N., Pentland, A. (Sandy), & Lazer, D. (2009). Inferring Friendship Network Structure by Using Mobile Phone Data. *Proceedings of the National Academy of Sciences*, 106(36), 15274–15278.

- Feiock, R. C. (2013). The Institutional Collective Action Framework. *Policy Studies Journal*, 41(3), 397–425.
- Feldman, M. W., Aoki, K., & Kumm, J. (1996). Individual Versus Social Learning: Evolutionary Analysis in a Fluctuating Environment. *Anthropological Science*, 104(3), 209–231.
- Friedkin, N. E. (2001). Norm Formation in Social Influence Networks. *Social Networks*, 23, 167–189.
- Friedkin, N. E., & Johnsen, E. C. (1997). Social Positions in Influence Networks. *Social Networks*, 19(3), 209–222.
- Friedkin, N. E., & Johnsen, E. C. (2011). *Social Influence Network Theory: A Sociological Examination of Small Group Dynamics*. Cambridge; New York: Cambridge University Press.
- Gerber, E. R., Henry, A. D., & Lubell, M. (2013). Political Homophily and Collaboration in Regional Planning Networks. *American Journal of Political Science*, 57(3), 598–610.
- Gerlak, A. K., & Heikkila, T. (2011). Building a Theory of Learning in Collaboratives: Evidence from the Everglades Restoration Program. *Journal of Public Administration Research & Theory*, 21(4), 619–644.
- Girard, Y., Hett, F., & Schunk, D. (2015). How Individual Characteristics Shape the Structure of Social Networks. *Journal of Economic Behavior & Organization*, 115, 197–216.
- Henry, A. D. (2009). The Challenge of Learning for Sustainability: A Prolegomenon to Theory. *Human Ecology Review*, 16(2), 131–140.
- Henry, A. D. (2011). Ideology, Power, and the Structure of Policy Networks. *Policy Studies Journal*, 39(3), 361–383.
- Henry, A. D., Ingold, K., Nohrstedt, D., & Weible, C. M. (2014). Policy Change in Comparative Contexts: Applying the Advocacy Coalition Framework Outside of Western Europe and North America. *Journal of Comparative Policy Analysis: Research and Practice*, 16(4), 299–312.
- Henry, A. D., Prałat, P., & Zhang, C.-Q. (2011). Emergence of Segregation in Evolving Social Networks. *Proceedings of the National Academy of Sciences*, 108(21), 8605–8610.
- Henry, A. D., & Volla, B. (2014). Networks and the Challenge of Sustainable Development. *Annual Review of Environment and Resources*, 39(1), 583–610.
- Ingold, K., & Fischer, M. (2013). Drivers of Collaboration to Mitigate Climate Change: An Illustration of Swiss Climate Policy Over 15 Years. *Global Environmental Change*, 24, 88–98.
- Innes, J. M. (1978). Selective Exposure as a Function of Dogmatism and Incentive. *The Journal of Social Psychology*, 106(2), 261–265.
- Jackson, M. O. (2008). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.
- Kandel, D. B. (1978). Homophily, Selection, and Socialization in Adolescent Friendships. *American Journal of Sociology*, 84(2), 427–436.
- Kleinbaum, A. M., Stuart, T. E., & Tushman, M. L. (2013). Discretion Within Constraint: Homophily and Structure in a Formal Organization. *Organization Science*, 24(5), 1316–1336.
- Kossinets, G., & Watts, D. J. (2009). Origins of Homophily in an Evolving Social Network. *American Journal of Sociology*, 115(2), 405–450.

- Krackhardt, D. (1999). The Ties That Torture: Simmelian Tie Analysis in Organizations. *Research in the Sociology of Organizations*, 16, 183–210.
- Lazer, D. (2001). The Co-Evolution of Individual and Network. *The Journal of Mathematical Sociology*, 25(1), 69–108.
- Leach, W. D., & Sabatier, P. A. (2005). To Trust an Adversary: Integrating Rational and Psychological Models of Collaborative Policymaking. *American Political Science Review*, 99(04), 491–503.
- Lewis, K., Gonzalez, M., & Kaufman, J. (2012). Social Selection and Peer Influence in an Online Social Network. *Proceedings of the National Academy of Sciences*, 109(1), 68–72.
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109.
- Marsden, P. V., & Friedkin, N. E. (1993). Network Studies of Social Influence. *Sociological Methods & Research*, 22(1), 127–151.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27, 415–444.
- Munro, G. D., & Ditto, P. H. (1997). Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information. *Personality and Social Psychology Bulletin*, 23(6), 636–653.
- Munro, G. D., Ditto, P. H., Lockhart, L. K., Fagerlin, A., Gready, M., & Peterson, E. (2002). Biased Assimilation of Sociopolitical Arguments: Evaluating the 1996 U.S. Presidential Debate. *Basic and Applied Social Psychology*, 24(1), 15–26.
- Noel, H., & Nyhan, B. (2011). The “Unfriending” Problem: The Consequences of Homophily in Friendship Retention for Causal Estimates of Social Influence. *Social Networks*, 33(3), 211–218.
- Olsson, A., & Phelps, E. A. (2007). Social Learning of Fear. *Nature Neuroscience*, 10(9), 1095–1102.
- O’Neill, H. M., Poudier, R. W., & Buchholtz, A. K. (1998). Patterns in the Diffusion of Strategies across Organizations: Insights from the Innovation Diffusion Literature. *The Academy of Management Review*, 23(1), 98–114.
- Parson, E. A., & Clark, W. C. (2005). Sustainable Development as Social Learning: Theoretical Perspectives and Practical Challenges for the Design of a Research Program. In *Barriers and Bridges to the Renewal of Ecosystems and Institutions* (pp. 428–460). New York: Columbia University Press.
- Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., ... Laland, K. N. (2010). Why Copy Others? Insights from the Social Learning Strategies Tournament. *Science*, 328(5975), 208–213.
- Sabatier, P. A., & Jenkins-Smith, H. C. (Eds.). (1993). *Policy Change and Learning: An Advocacy Coalition Approach*. Boulder, Colo.: Westview Press.
- Sandberg, J., Rytina, S., Delaunay, V., & Marra, A. S. (2012). Social Learning About Levels of Perinatal and Infant Mortality in Niakhar, Senegal. *Social Networks*, 34(2), 264–274.
- Saramaki, J., Leicht, E. A., Lopez, E., Roberts, S. G. B., Reed-Tsochas, F., & Dunbar, R. I. M. (2014). Persistence of Social Signatures in Human Communication. *Proceedings of the National Academy of Sciences*, 111(3), 942–947.

- Schelling, T. C. (1971). Dynamic Models of Segregation. *The Journal of Mathematical Sociology*, 1(2), 143–186.
- Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New Specifications for Exponential Random Graph Models. *Sociological Methodology*, 36, 99–153.
- Stehlé, J., Charbonnier, F., Picard, T., Cattuto, C., & Barrat, A. (2013). Gender Homophily from Spatial Behavior in a Primary School: A Sociometric Study. *Social Networks*, 35(4), 604–613.
- Weible, C. M., & Sabatier, P. A. (2005). Comparing Policy Networks: Marine Protected Areas in California. *Policy Studies Journal*, 33(2), 181–201.
- Weible, C. M., Sabatier, P. A., Jenkins-Smith, H. C., Nohrstedt, D., Henry, A. D., & deLeon, P. (2011). A Quarter Century of the Advocacy Coalition Framework: An Introduction to the Special Issue. *Policy Studies Journal*, 39(3), 349–360.

Figure 1: A framework for analysis

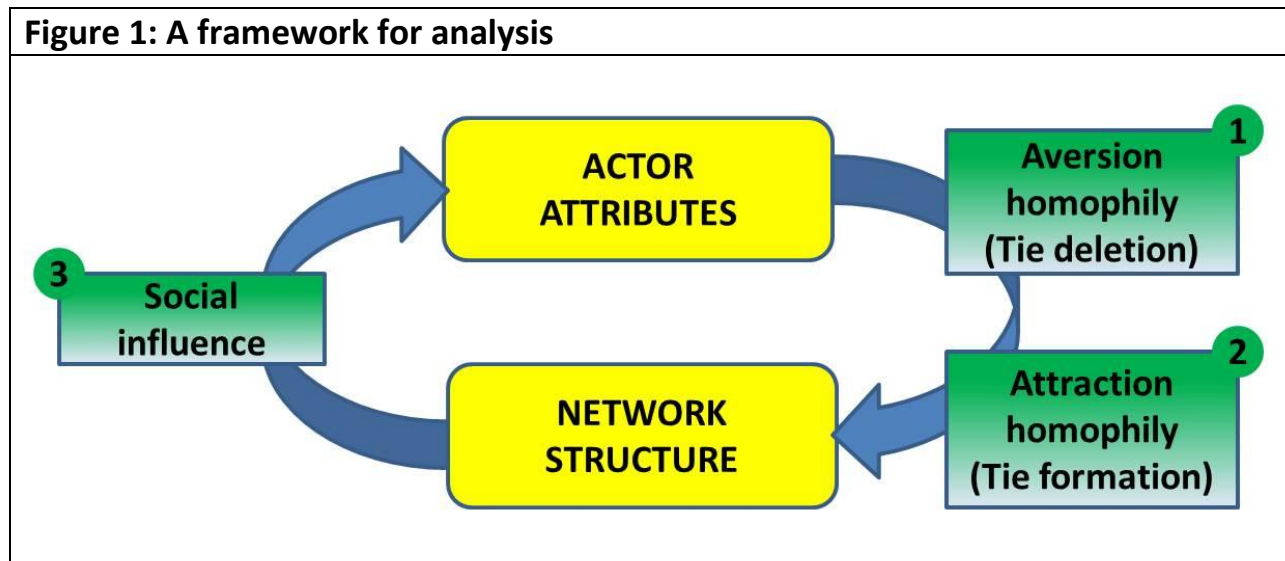


Table 1: Descriptive statistics for two network snapshots

	October 19th	November 22nd
Total number of nodes	3.9 million	4.3 million
Total number of edges	6.1 million	6.9 million
Average node degree	3.15	3.15
Size of connected component	3.3 million (85%)	3.7 million (85%)
Total number of components	145,000	172,000

Figure 2: Cumulative degree distributions of two network snapshots

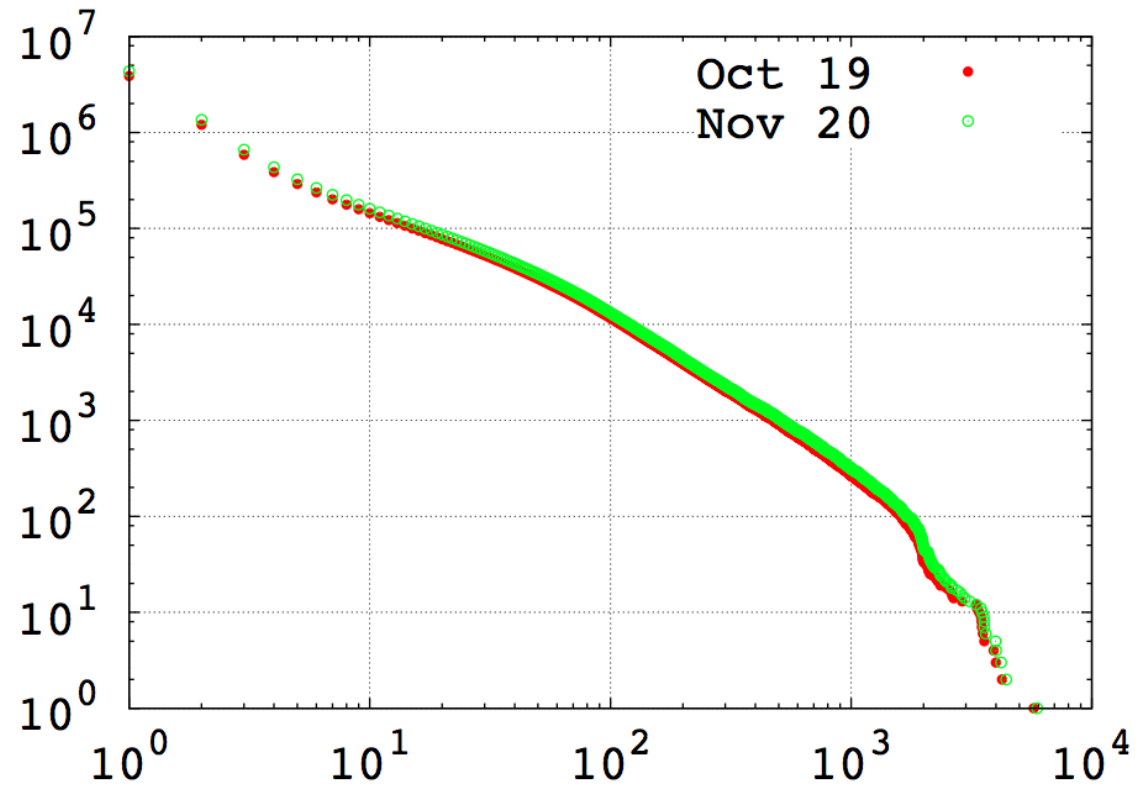


Table 2: Four categories of dyads

		<i>In the second time period...</i>	
		<i>...tie exists</i>	<i>...tie does not exist</i>
<i>In the first time period...</i>	<i>...tie exists</i>	Stable tie (Group 1-1)	Deleted tie (Group 1-0)
	<i>...tie does not exist</i>	Newly formed tie (Group 0-1)	Stable non-tie (Group 0-0)

Figure 3: Distributions of geodesic path lengths in first time period (October 19th) between disconnected actors who later formed a tie (orange) versus those who did not form a tie (white)

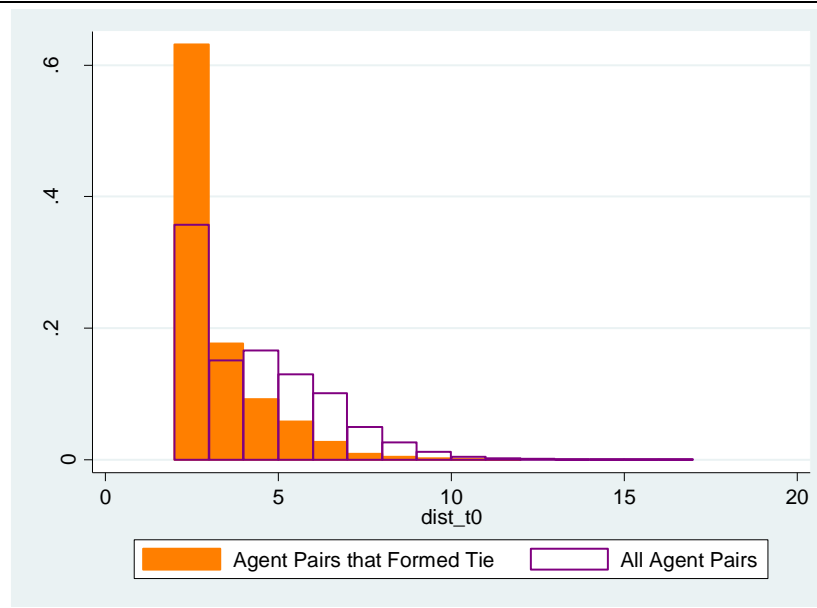


Table 3: The effect of attribute differences on tie formation and tie deletion

	Model 1 (DV=tie formation)	Model 2 (DV=tie formation)	Model 3 (DV=tie deletion)	Model 4 (DV= tie deletion)
Euclidean distance	-0.135 (0.041)		0.176 (0.032)	
Hamming distance		-0.053 (0.016)		0.072 (0.013)
Path length in first time period	-1.056 (0.035)	-1.056 (0.035)		
Constant	4.048 (0.164)	3.998 (0.156)	-0.873 (0.085)	-0.820 (0.078)
N	3,540	3,540	3,328	3,328
pseudo-R ²	0.2767	0.2766	0.0070	0.0068

Note: Standard errors reported in parenthesis.

Figure 4: Effect of attribute differences on the probability of tie deletion

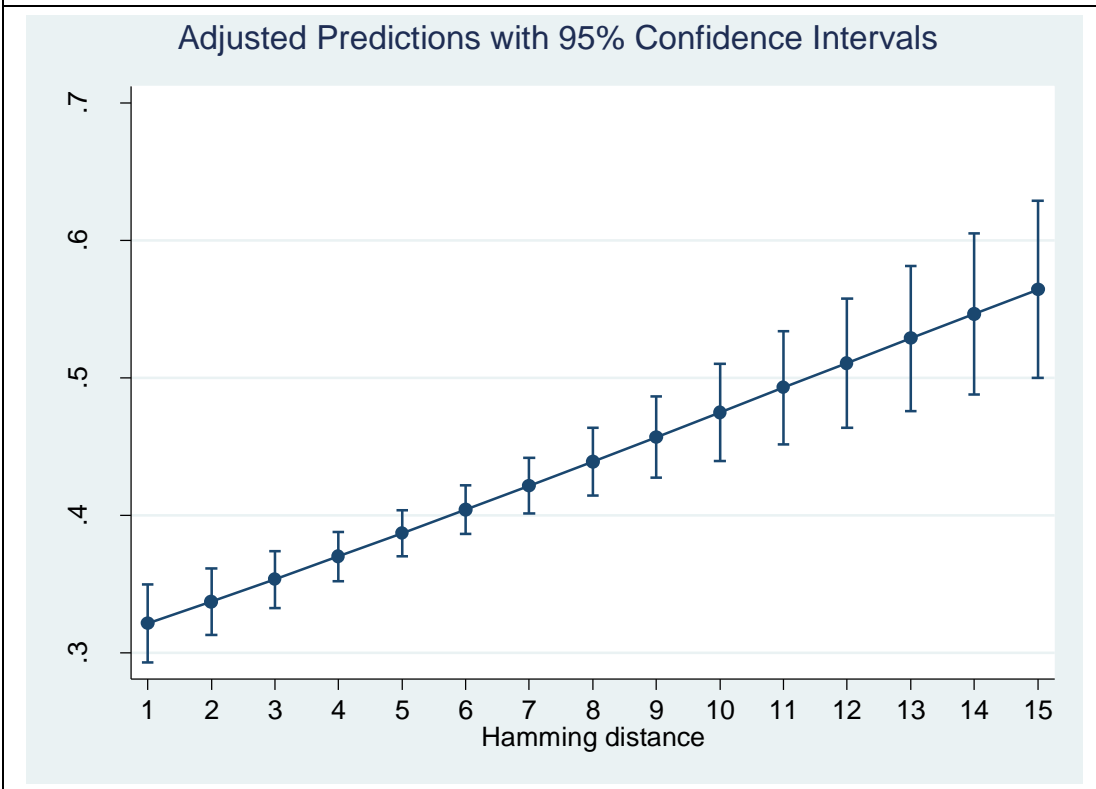
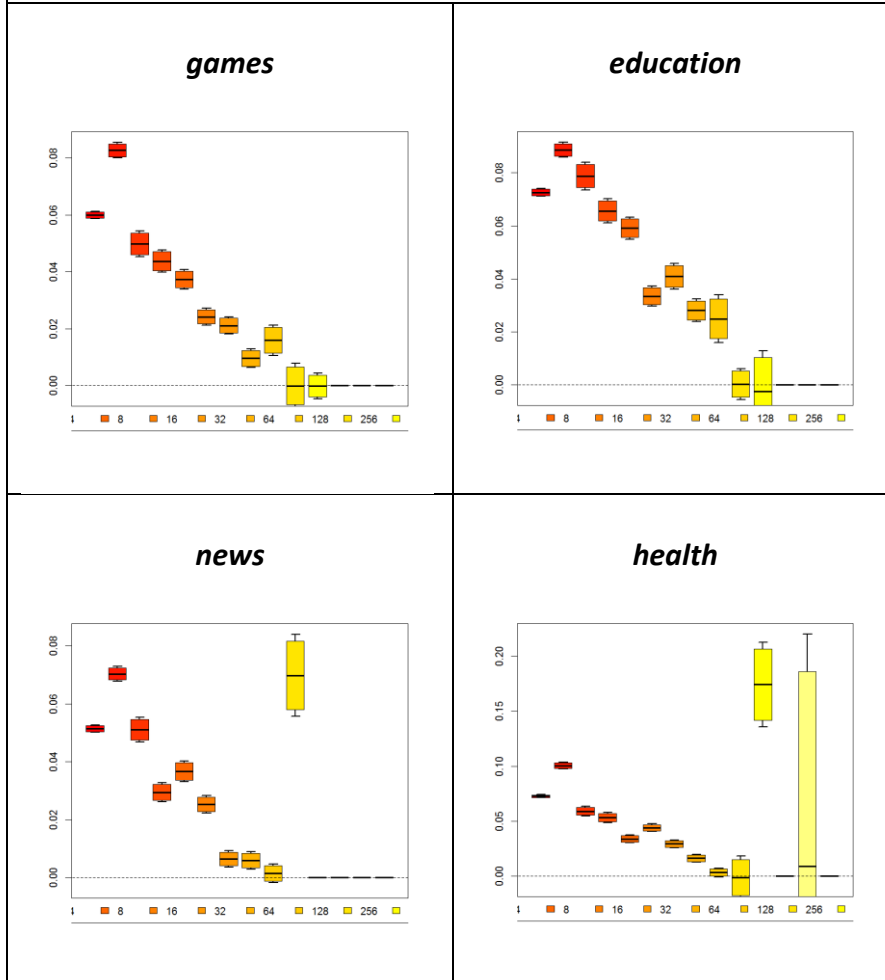


Table 4: Estimated influences of global and local centers of mass on user tastes

<i>Model (taste element)</i>	<i>Effect of change in global center of mass (b)</i>	<i>Effect of change in local center of mass (a)</i>	<i>R²</i>
lifestyle	0.223 (0.012)	0.768 (0.002)	0.58
finance	0.897 (0.044)	0.077 (0.001)	0.06
business	0.870 (0.037)	0.070 (0.001)	0.05
entertainment	0.869 (0.032)	0.079 (0.001)	0.06
ringtones	0.955 (0.258)	0.019 (0.000)	0.02
photo	0.860 (0.031)	0.085 (0.001)	0.06
news	0.908 (0.041)	0.051 (0.001)	0.04
utilities	0.792 (0.023)	0.154 (0.001)	0.12
health	0.904 (0.046)	0.073 (0.001)	0.06
sports	0.903 (0.051)	0.050 (0.001)	0.04
themes	0.976 (0.068)	0.017 (0.000)	0.01
books	0.890 (0.061)	0.060 (0.001)	0.04
games	0.908 (0.032)	0.060 (0.001)	0.05
social	0.445 (0.014)	0.545 (0.002)	0.43
productivity	0.664 (0.022)	0.287 (0.001)	0.22
education	0.877 (0.041)	0.073 (0.001)	0.05
navigation	0.785 (0.026)	0.159 (0.001)	0.12

Note: Standard errors are reported in parenthesis.

Figure 5: Distribution of changes in user tastes as a function of number of social contacts



Note: Boxplots represent distributions of individual user shifts in tastes between October 19th, 2013 and November 22nd, 2013. Distributions are conditional on degree of user on October 19th. Boxplots exclude outliers.