Network Propagation of Connected Ideas

James Houghton
houghton@mit.edu

Proceedings of the 35th International Conference of the System Dynamics Society
Cambridge, Massachusetts, USA July 16 – July 20, 2017

Abstract

In his classic work on collective behavior, Granovetter explicitly excludes the question of 'how individuals happen to have the preferences they do' for participation in social contagion. This paper suggests that the individual characteristics driving adoption of ideas are themselves a product of social influence, and that by allowing that ideas have meaning only in relationship to one another, this constructive process can be fruitfully explored. This paper presents a highly-simplified formalization of relationships between ideas diffusing on a social network to show how semantic network edge propagation may be responsible for the macro-level phenomena of factionalism in well-connected graphs under simple contagion.

Introduction

Theories of idea propagation in social networks presently struggle to reconcile individual susceptibility to adoption with the reality that adoption influences future susceptibility to new ideas. Models of probabilistic or threshold contagion give insight into the way network structures such as weak ties, structural bridges, and clustering influence the extent and rate of propagation. At the individual level, both theoretical and empirical studies show homophily and assortativity on individual characteristics to account for significant components of diffusion. However, the way these individual characteristics emerge in the context of social networks remains largely unexplored.

In this paper I will suggest that the individual characteristics driving adoption are themselves a product of social influence, and that by allowing that ideas have meaning only in relationship to one another, this constructive process can be fruitfully explored. Characteristics of susceptibility are determined by the set of ideas that individuals hold, and these characteristics go on to shape the ideas individuals are exposed to and adopt through social contagion.

Existing theories of idea propagation are largely concerned with individual access to novel information, from an individual benefit or marketing perspective. This follows the interest of seminal works such as Granovetter's theory of weak ties (1973) and Burt's arguments about structural holes and brokerage (1992); or Kempe, Kleinberg and Tardos's (2003) optimization of social influence interventions. Empirical research which investigates both meaning and structure has only recently been enabled by modern data science, and studies such as those by Uzzi (1997) or Aral and Alstyne (2011) continue to evaluate and extend theory of how network structure and propagation processes influence access to information.

The reverse process, by which meaning influences propagation, lacks a theoretical grounding that is compatible with network analysis and corresponds to empirically observable phenomena. In his other classic work on threshold models of collective behavior, Granovetter (1978) explicitly excludes the question of 'how individuals happen to have the preferences they do' for participation in social contagion. This paper suggests in the style of Granovetter a simple theoretical model whose emergent behavior includes the process in question.

This paper explores one highly-simplified representation of the relationship between ideas to show how semantics may be responsible for the macro-level phenomena of factionalism in well-connected graphs under simple contagion. While the present offering is largely theoretical, a reader familiar with natural language processing and co-occurrence analysis will recognize the opportunity for empirical evaluation in social media contexts.

Semantic Networks

It is hardly novel to say that ideas are related to one another. Some ideas share a connection of similarity or ontology, others of subject and object, still others of syntax or language. Connections may be strong or weak, binding or exclusive, positive or negative. The implications of each of these dimensions for theoretical and empirical research will deferred to future work. For the purposes of this paper it is sufficient that a connection exists or does not exist between two ideas.

These connections, however, are not a universal, objective reality. They are instead subjectively held within the minds of individuals. While there are a number of ways this could be conceptualized and formalized, I choose to represent a connection between ideas as part of a semantic network unique to each individual, as seen in Figure 1. The semantic network contains a node for each idea, and an edge between ideas that are in some way related to one another. Individuals may make new connections between existing nodes, or add nodes through their connection with existing or other new nodes.

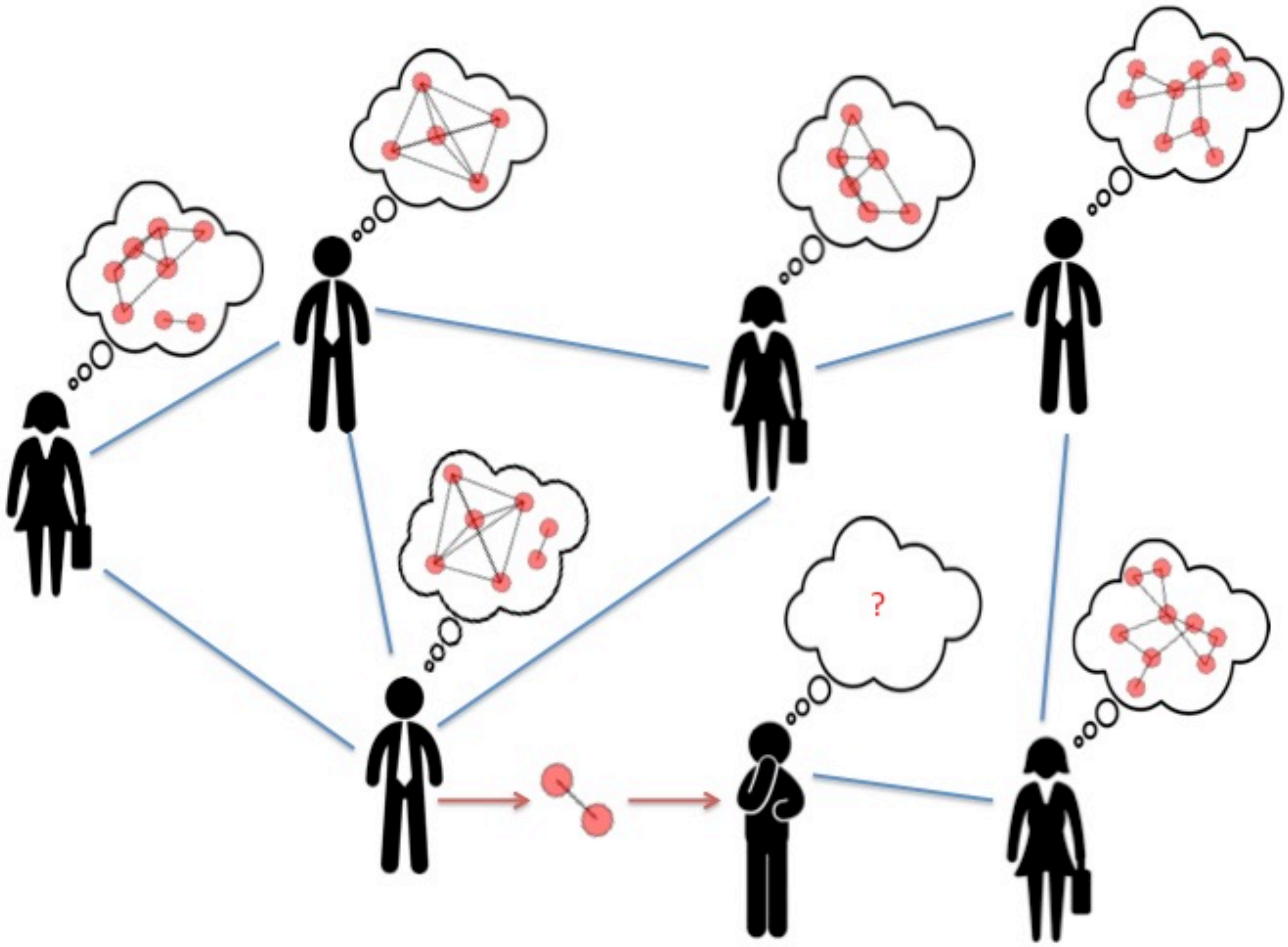

Figure 1: Semantic networks present within the minds of individuals in a social network

The first hypothesis presented to link idea propagation with its influence on future susceptibility is that connections between concepts are the relevant entities of idea propagation on a social network, and it is these *connections* which may be adopted or ignored by individuals. This hypothesis abandons the traditional formalization of ideas as unitary, independent entities in their propagation and adoption. Instead, the formalization requires that ideas have context, and in order to mean something to their recipients must be spread concomitantly with that context.

This hypothesis could be argued by suggesting that ideas only have meaning through their relationships to other concepts, as without those relationships words are merely sound, images are merely pixels. More pragmatically, however, this hypothesis is motivated by examples such as the process of learning language: the word *Adroit* is synonymous with the word *Masterful*; or by how introductions are made at parties: *Freida* is a *Formal Modeler*. Depending on the party, the connections made in such an introduction may spark discussion by relating something new to something known, or may leave the interlocutor stymied.

The internal consistency of this abstraction is supported by the use of semantic networks for knowledge representation in linguistic and artificial intelligence tasks (Woods 1975, Sowa 1991, Speer and Havasi 2012). These methods likewise represent concepts as nodes in a network, and use edges to represent relations between these concepts. In many situations these edges are categorized by relationship types, and the networks are organized into taxonomic hierarchies in which properties are inherited from more general categories.

Evidence for external validity within human cognitive processes comes from studies of semantic memory exemplified by Collins and Quillian (1969). These studies use the time taken for retrieval of information requiring various numbers of 'hops' in a conceptual network as evidence that the information being retrieved is indeed stored through a chain of connections, and not as a set of isolated facts. Tulving (1972), and Posner et. al. (1988) go on to explore more deeply the physical mechanisms and conditions for storage and processing of information through semantic networks in the brain.

In recent years, machine learning algorithms using frequency methods trained on large datasets have begun to supplant interpretable semantic networks for natural language tasks. Similarly, research into the biophysiology of semantic representation has shifted away from investigating semantic networks in vivo. These trends have limited the support for literal interpretability of the semantic network abstraction. However, the implications of this hypothesis in the context of the present research suggest methods of empirical verification that may support the argument.

For the purposes of this paper, it is not sufficient to note that connections between concepts form an individual's notion of meaning. A mechanism must also be described by which the cross-individual differences in semantic networks account for variances in susceptibility to new contagion. Likewise must be described how the structure of these connections within an individual makes them susceptible to adopting one connection over another.

As with the prior hypothesis, there are many ways in which this influence could be conceived and formalized. I will choose a simple heuristic, suggesting that individuals prefer to adopt edges which join nodes that are already 'close' to one another in their semantic network. Concepts of network closeness may include various measures of connectivity, centrality and path redundancy. Exploration of these alternatives is left to future research, as for the purposes of the present exploration it is sufficient to consider the minimum path length between two nodes, should one exist, or alternately a heuristic measure that favors minimum change to the existing network.

For example, consider the semantic network illustrated in Figure 2. Red nodes (0 - 5) and solid lines exist within the mental network, and dashed lines are candidates for adoption. Edge (0, 2) is most likely to be adopted, as the nodes it joins are closely connected via another route. Edge (0, 3) follows in preference, as the nodes it connects are further apart in the existing network. Edge (3, 4) is the third most likely, as it connects otherwise disjoint existing nodes. Edge (5, 6) is the next most likely, as one of its end points exists in the individual's semantic network. Edge (7, 8) is the least likely to be adopted, as neither node exists within the network.

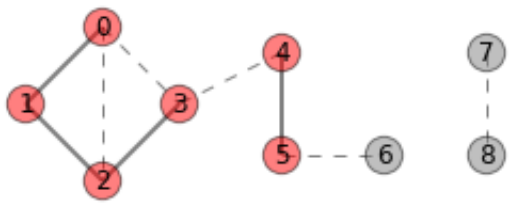

Figure 2: Edge adoption priority: (0, 2) > (0, 3) > (3, 4) > (5, 6) > (7, 8).

Thus, individuals have a preference for adopting edges that maximize the number of connections per node, and for creating mutually interrelated structures of meaning.

The hypothesized preference arises due to the cognitive load associated with making a connection between two ideas. Intuition suggests that it is easier to see how two concepts are related if one can quickly draw a path of inference between the two based upon existing knowledge. The most difficult task should be to learn new information that has no connection with any existing knowledge. This phenomenon is responsible for the fact that university courses have specific prerequisites - the ability to learn new material is not merely a function of the overall volume of knowledge a student possesses, but depends on the presence of a specific mental framework to which new ideas can be grafted. For a more concrete example, consider two statements that an individual might encounter: "Beethoven's Moonlight Sonata was originally written for the harpsichord" and "avocados are berries". Despite being false, the first statement makes a connection between a piano piece and the harpsichord instrument. As pianos are similar to harpsichords, the mental distance is small, and the statement seems plausible. The second (correct) statement requires the reader to consider berries as fruits containing seeds from a single flower, and due to this semantic distance may not 'hang together' as well at first blush.

Empirical support for this assumption comes from Linn (2004), "Learning involves generating connection among ideas… learners rarely generate connections across contexts… Experience, prior knowledge, and context make some connections salient and others invisible." Also in this vein is Schilling's (2005) model, which represents cognitive insight as a shortcut in a semantic network that bridges previously disconnected ideas. In Shilling's model, meaning is built as ideas come into closer relationship with one another. Connections built between ideas which are close to one another form the basis for incremental improvement in knowledge, whereas the rare 'Aha!' arises as connections are made between concepts with high network distances. Shilling suggests that under-connected structures in a cognitive network represent problems to be solved. Here see also Mayer (1995).

The degree of preference for minimizing change and closing semantic network structures is an open area for empirical study. By varying the preference for minimizing the structural change to semantic networks, the present research suggests consequences of different levels of preference that are empirically observable at the macro-level, and which may help build further confidence in the hypothesis.

Semantic Contagion

In order to simulate the consequences of the given hypotheses, they must first be explicitly formalized. To do so requires description of the learning and memory processes of the individual that drive their handling of concepts, the transactive processes that drive their relationships with their neighbors, and the social network structure over which these processes operate. A fully behavioral model might consider spontaneous generation of connections, ascribed characteristics of individuals, strategic behavior in communication, broadcast communication pathways, or endogenous link formation. For the purposes of this paper, however, I will assume a population of identical agents who communicate uniformly and exclusively with their neighbors over a fixed and well-connected network. With these assumptions, the emergent macro-level behavior may only emerge from the interaction between ideas in individuals' semantic networks. As will be shown as the model is explored, these structural assumptions have consequences for factionalism and information transfer that are only expected to strengthen when the assumptions are relaxed to include endogenous network formation and strategic communication. These extensions go a long way toward a formal theory of filter bubbles, although such exploration is deferred to future research.

The primary mechanisms of the formal model presented here are those responsible for communication, learning, and memory of edges in a semantic network. To make these processes clear, I will describe them in the context of a single interaction between friends. While the process is here presented asymmetrically, in fact each step occurs for each individual simultaneously.

Imagine two friends, Alice and Bob. Alice and Bob have similar, but nonidentical views of the world. In Figure 3.a, Bob's initial semantic network is shown as three edges connecting five nodes in two groups. [1]

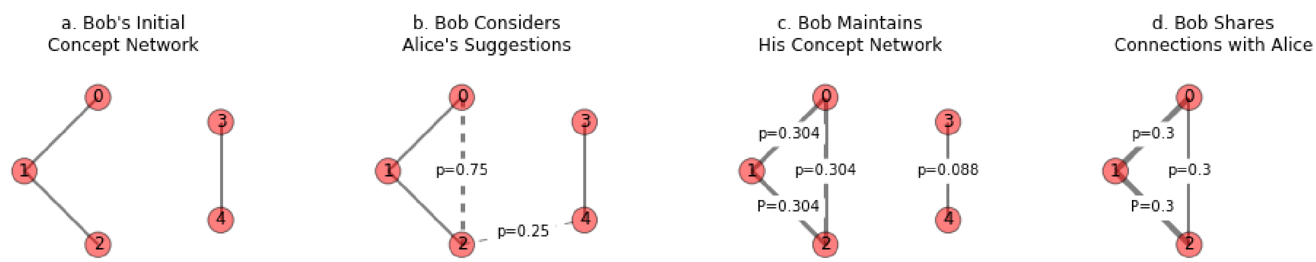

Figure 3: Bob's process

Alice emits from her own semantic network (not shown) edges (0, 2) and (2, 4). Bob would like to learn from Alice, but slow on the uptake, Bob only has the capacity to absorb one of the connections he has received. Due to his preference for making small changes to his semantic network Bob is more likely (but not guaranteed) to adopt edge (0, 2) and ignore edge (2, 4). While for Bob this is an implicit subconscious negotiation happening between groups of neurons, it's simulated representation must be more explicit. To model Bob's decision process, assume that the likelihood of an edge being adopted is proportional to a score representing the inverse distance that would be spanned by the edge, raised to the power of bob's preference for making small changes to his mental network:

$$\left(\frac{1}{Length_{(\alpha,\beta)}}\right)^{Preference}$$

Equation 1

This formulation has the advantage that if Bob's preference is zero (that is, he has no preference for connected mental maps) then there will be equal likelihood of all edges being adopted. If the preference is very strong, then the stochasticity will be suppressed, and the 'shortest' links will always be made. If the preference is negative, Bob can explore the consequences of intentionally seeking out unusual information. Assuming that Bob's preference takes the value 1, a score and normalized probability for adopting each of these edges can be evaluated as in Table 1:

Table 1

Edge	Length	Score	Probability
(0, 2)	2	$(½)^1$	0.75
(2, 4)	6	$(1/6)^1$	0.25

What sticks out in this table is the length of edge (2, 4). As there is no existing connection between the two nodes, this formulation uses one more than the number of nodes in the graph as a heuristic value of the length. To ensure that the prioritization scheme is consistent with that described in Figure 2, double this for each missing node. While there are many ways this scoring system may be implemented, the above is simple and sufficient for the present purposes.[2] In one realization of this stochastic process, Bob adopts edge (0, 2).

In addition to his finite capacity to absorb new knowledge each round of the simulation, Bob has a finite memory capacity. He can only maintain so many edges in his semantic network at a time - sadly for Bob, that is three. Having adopted a new edge from his interaction with Alice, Bob now forgets an edge, with similar propensity to remember edges that participate in well-connected structures. To simulate the process of forgetting, or conversely of maintaining edges in memory, it is sufficient to recompute the scores and probabilities for each edge, assuming it were not already present. A similar preference parameter is present - in this case, a preference for remembering well connected edges. For the sake of explanation, Bob has been given a preference of 0.5 for maintaining connected edges. Table 2 enumerates the results for Bob's semantic network as shown in Figure 3.c.

Table 2

Edge	Length	Score	Probability
(0, 1)	2	$(½)^{0.5}$	.304
(1, 2)	2	$(½)^{0.5}$	.304
(2, 0)	2	$(½)^{0.5}$	.304
(3, 4)	24	$(1/24)^{0.5}$	.088

Note that the length assigned to edge (3, 4) is the no-path length (5 nodes + 1) penalized (doubled) by the number of nodes that would be missing from the graph if the edge were not present, that is, both of them. The simulation of Bob's mental process chooses without replacement edges to maintain to the next round up until the limits of Bob's memory capacity. [3] In one realization of the stochastic process, Bob maintains the three connected edges, and forgets edge (3, 4).

From his new understanding, Bob is now ready to share his thoughts with Alice. I leave notions of preferential or strategic emission to future research and suggest that Bob emits randomly from his network, up to his finite emission capability, as shown in Figure 3.d. In one particular realization of the stochastic process, Bob emits edges (0, 1) and (1, 2), which Alice will receive at the start of the next round.

This example demonstrates the simulation of a single agent's learning process. It is straightforward to expand the simulation to a full society. In addition to receiving emissions from Alice, each round Bob also receives emissions from neighbors Carol and Dave, and emits uniformly to all three. Alice may have different friends from Bob, with whom she interacts. Thus, an edge may propagate through a network to the extent that it is preferentially selected in the process of updating existing, socially connected semantic networks.

Even this qualitative description gives some insight into the three problems mentioned in the introduction. Factionalism may be considered to be the situation in which groups of individuals have highly similar semantic networks, with significant differences across groups. Partial contagion may occur when an edge propagates well within a faction due to the similarity between semantic networks within that group, and then stops at the faction boundary because the infrastructure necessary to support adoption is absent. Heteromorphous diffusion processes may occur if a single individual emits first an edge that is well received by his faction, and then a second which does not fit comfortably into that faction's semantic network, but is well adopted by another.

Measuring Factionalism

From an observational point of view, an individual should conclude that 'factionalism' was present in situations in which he found a significant number of individuals who shared with him his worldview, and another significant number of individuals who, while disagreeing with him, agreed with one another. From the global perspective, this translates to the observation that disagreements across factions should be larger than disagreements within factions. Omitted from this definition are any concepts of network structure or any of its proxies such as organizational membership, leadership or formal agendas. This is consistent with the desire to show how factionalism can emerge independently from network structure. Also omitted are notions of the number of factions that should exist, or the size of the factions. Likewise omitted is any notion of the average level of agreement of the population. Factionalism may occur in situations in which the average agreement between individuals is 10% or 90%, just so long as disagreement across groups exceeds disagreement within groups. Factions may be nested, such that within the realm of differences within a group, subpopulations may form who are even more self-similar, in opposition to other members of the larger faction.

The proposed definition is consistent with concepts used for machine learning clustering of data: "clustering algorithms seek to construct clusters of records such that the between-cluster variation is large compared to the within-cluster variation" (Larose and Larose 2014.) Hierarchical clustering will form the basis for a continuous metric for factionalism. In hierarchical clustering, the two most similar elements are combined recursively until a tree is constructed based upon distances between the various components, as shown in the dendrograms plotted in Figure 4.

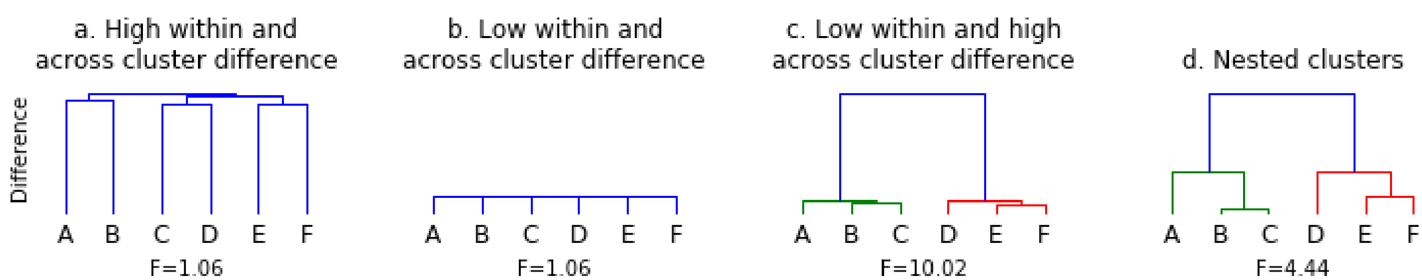

Figure 4: Dendrograms illustrating the presence of clusters

Figure 4 highlights various possibilities for clustering. In this diagram, individuals are paired with the individuals they are most similar to, and then similar groups are recursively paired until the entire population is hierarchically clustered. The heights of each horizontal bar indicates the level of disagreement between the two groups it bridges. The diagram allows inspection of both the overall level of agreement, and the possibility of factionalism within the population. In 4.a, there is very little similarity between Alice, Bob, or any of their contemporaries. The groups that form have high levels of disagreement within the cluster (as indicated by the high levels of the bars between (A, B), etc. which form the clusters. The level of disagreement across clusters, as indicated by the bars which span the (C, D) and (E,F) groups is barely higher than the level of disagreement within the groups, and so no factionalism is evident. Likewise in Figure 4.b, there is in this case low levels of disagreement within the groups, and again low levels of disagreement across the groups. Figure 4.c shows the quintessential clustering pattern of low disagreement within groups (A, B, C) and (D, E, F) and high levels of disagreement across groups. Figure 4.d reveals that clustering is still possible with moderate levels of within-group disagreement, so long as the difference between groups is still significantly larger than the differences within groups.

These charts motivate a metric for clustering and factionalism within the population, which is the maximum level of disagreement in the hierarchical clustering set (corresponding to the hight of the bar which finally connects the last remaining subgroups) normalized by the average height of the remaining bars. In Figure 4, this metric labeled 'F' shows identical (low) values in the first two cases, and greater values in subfigures c and d. The factionalism metric F can be interpreted approximately as the ratio of disagreement across factions to the disagreement within factions when there are two primary factions shown. Note that this is not a strict comparison of all intersubjective pairs, but is based upon the hierarchical clustering algorithm. This serves to weight two strong factions as being more 'factionalized' than three. This is partly an aesthetic choice, as two-party conflict seems to be more frequently observed than that which is truly multiparty. It is also partly a computational choice - this metric is faster to compute than full factorial comparison.

It now remains to determine what is meant by 'agreement' and 'disagreement'. If Alice and Bob have identical semantic networks, clearly they should have 100% agreement. But if each individual's network contains 10 unique edges, with no overlap, how should their agreement be assessed? There are an enormous number of edges which neither maintains - should these contribute to their similarity? Not likely. It is not sufficient then to compare a binary feature vector over all possible edges. Instead, define a metric of similarity using the Jaccard Coefficient, the intersection of the sets of edges in the semantic network of each, as a fraction of the union of these sets. Thus in the first case above, agreement is equal to one, and in the second case it is equal to zero.

$Similarity_{A,B} = \frac{A \cap B}{A \cup B}$ **Equation 2**

As this metric is simply defined over the 0-1 interval, a natural interpretation of 'disagreement' is that it is simply $1 - Similarity_{A,B}$, or the Jaccard Distance.

Emergent Factionalism

One of the earliest and most fundamental observations of human social networks is that they exhibit 'small world' characteristics (Travers and Milgram, 1969, and others). These networks are typified by short average path lengths between randomly chosen pairs of individuals, and high network connectivity. Watts and Strogatz (1998) showed that these types of networks are highly conducive to social contagion, building on Granovetter's (1973) insight that few, long distance 'weak ties' help to transfer information between clusters. Watts (2002) showed that when networks are highly clustered, the extent of a contagion cascade is bimodal – either spreading through (approximately) the entire population, or being confined to the originating cluster.

If these observations are true, how do factions form? What mechanisms account for the observation that ideas may spread easily through a group of individuals but fail to cross faction boundaries, even when these factions are otherwise well connected in the social network?

Existing literature suggests three approaches to addressing this puzzle. Zachary (1977) introduces a concept of social network tie strength, suggesting that factionalism can exist in well connected networks when tie strength between factions was less than that within factions. His empirical results show that factions can be identified prior to group fission based on the structural characteristics of the network, but he does not suggest how these tie strengths are formed, or specify an endogenous theory of factionalism. Centola and Macy (2007) introduce the idea of complex contagion, and demonstrate that when multiple sources of influence must be present for adoption, incomplete propagation is possible. The extent of this form of super-linear influence weighting is to date under-explored, but it is unlikely to be the sole mechanism responsible for incomplete diffusion. This is particularly true of relatively costless opinion adoption, a prototypical example of simple contagion. A final approach typified by Aral et al. (2009) and Shalizi and Thomas (2011) is to suggest that incomplete diffusion is due to exogenous factors which both make individuals susceptible to adoption and likely to be connected with other individuals that share their susceptibility. Incomplete diffusion is thus a product not of social factors so much as demographic factors. While the empirical work associated with this research shows that homophily can account for a significant fraction of diffusion, ample room remains for true social influence. Additionally, ascribed characteristics remain to be distinguished from characteristics which are acquired through social processes.

To gain some intuition for how macro-level social structures may develop, it is instructive to consider the emergence of factionalism from a purely random initial state, in which no *a priori* structure is assumed whatsoever. In Figure 5, a randomly connected network of 40 individuals is instantiated, with average degree 3.8 and network clustering coefficient 0.04. Each agent is given a random sample of 10 edges for their semantic network, with potential nodes valued between 1 and 20, and thus a selection of up to 400 edges. By chance, 170 of these edges are included, and thus some overlap exists. In each round of the simulation, agents emit 2 of the edges from their semantic networks, update up to 4 edges in their semantic network based upon their neighbor's emissions and their preferences for semantic closure, and finally maintain their semantic network with a maximum of 10 edges.

The first column of images in the figure describes the state of the system at t=0. The dendrogram reveals high within-cluster and across-cluster disagreement between individuals and a Factionalism index slightly above 1. The second row shows the similarity between individuals. The similarity of an individual to each other member of the society is shown in each row and column. Rows and columns are sorted identically to bring similar individuals together, and so groups of likeminded individuals will appear as redder areas in the heat-map. The diagonal represents self-similarity, which is always 1. Below the time-series charts is a snapshot of the social network, with individuals colored by faction. At t=0, no factions exist, and so all members are colored together. The last row of images shows a sample semantic network from a particular individual (Alice, for the sake of discussion). At t=0. Alice's semantic network contains randomly selected edges. Some of the edges link to form chains, but little structure is evident.

As time progresses, the initial flurry of changes to every semantic network declines. Both the level of factionalism and the average similarity between individuals grows as semantic structures begin to form and reinforce one another. At t=500, the beginnings of clusters of similarity are evident within the dendrogram and similarity matrix. Nested clusters are identified and their members are colored within the social network graph.

Alice's semantic network is more structured, as disconnected edges are pruned in favor of those that participate in relationships with other concepts. At t=1000, clusters have begun to consolidate, and a signifiant portion of the population share identical mental structures. The number of clusters identified has shrunk, and the remaining clusters have more members. At t=1500, clusters have merged into two clearly defined groups. The factionalism index, the ratio of of the disagreement across the largest cluster grouping to the average disagreement within clusters, exceeds 10, largely on the strength of the large population of perfectly likeminded individuals. On average about 5 edges are changed anywhere in the population in the final rounds, implying that mental structures are largely stable. The average similarity between any two individuals exceeds 50%, again due to the large group of homogenous individuals present.

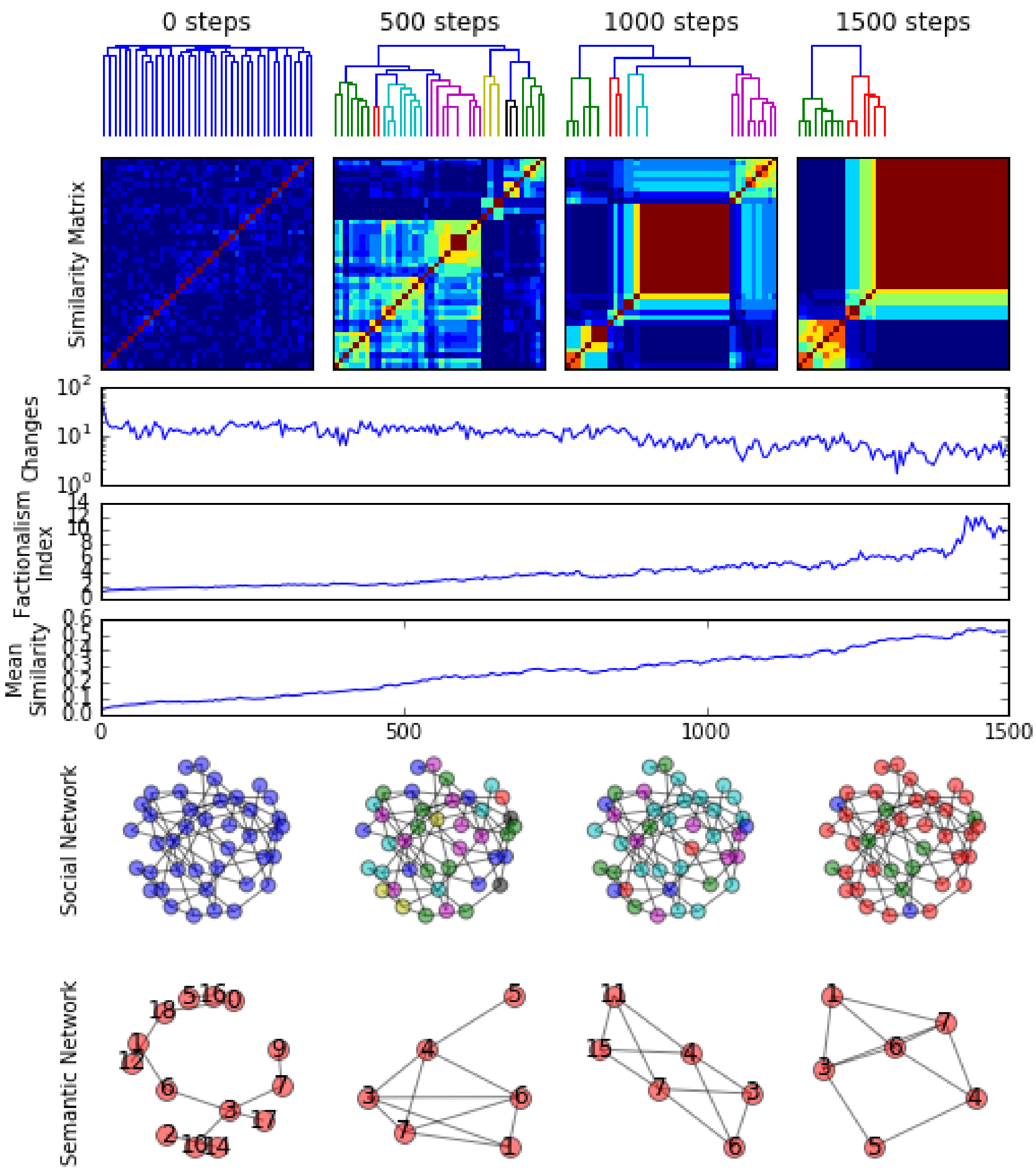

Figure 5: Faction development

Overcoming factionalism

Having elaborated a theory for how factionalism develops, it would be interesting to explore strategies for how factionalism may be overcome. That is, in a situation that is heavily factionalized, what can be done to bring individuals together? There are two basic strategies to consider. The first involves making social connections between individuals from different factions. A second strategy involves strategically introducing new edges into the mix of semantic networks. It may be the case that some combination of structural rewiring and new edge development performs better than either of these policies in isolation. This integrated strategy may be exactly what occurs in the real world when actors working for reconciliation (such as legal mediators, or programs such as MIT's MEET which brings Israeli and Palestinian youth together around a shared task) attempt to achieve understanding between parties.

There are likely to be objectively preferred intervention strategies in these situations. One might seek an optimal selection of social network connections to build, and an optimal set of themes for discussion to best bridge the semantic gap between factions and work for reconciliation between larger populations. Just as the marketing literature has explored optimal strategies for influence in social networks, there is room for those seeking reconciliation and working for peace to identify intervention strategies optimized through knowledge of social and semantic network structure.

Summary

This paper presents a theoretical model of factionalism, based upon the social construction of shared meaning structures, which accounts for individual level susceptibility to adopting ideas as an endogenous process of past contagion. The model is valuable in helping to understand varying morphologies of diffusion in situations where the underlying social network structure does not place intuitive limits on contagion. These situations may occur more often than not, as studies of homophily in social contagion reveal that individual characteristics constitute significant drivers of diffusion. The success of prior literature in using social structure and individual characteristics to explain limits to diffusion has led researchers to overlook the fact that these structures and individual characteristics are themselves often formed in response to social interaction.

By explaining how groups of individuals come to share similar meaning structures, semantic edge propagation models take the burden for explaining limits to diffusion away from structural models or reliance on exogenous factors and return it to the contagion process itself. These models may be useful in studies of factionalism within organizations as well as at the population level. The promise of this approach is to make available to analysis the mechanisms by which factionalism and susceptibility emerge, so that policies for intervention may be cogently crafted. The need for sociologically grounded policies for building bridges between factions both within organizations and across society is evident. Developing tools and theory for understanding the relevant social dynamics and evaluating potential interventions is one contribution to this important work of reconciliation.

References

Aral, S., & van Alstyne, M. (2011). The Diversity-Bandwidth Trade-off. American Journal of Sociology, 117(1), 90–171.

Aral, S., Muchnik, L. E. V, & Sundararajan, A. (2013). Engineering social contagions: Optimal network seeding in the presence of homophily. Network Science, (2013), 125–153. https://doi.org/10.1017/nws.2013.6

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. https://doi.org/10.1088/1742-5468/2008/10/P10008

Burt, R. (1992). The Social Structure of Competition. Networks and Organizations, 57–91.

Caliński, T., & Harabasz, J. (1974). “A dendrite method for cluster analysis”. Communications in Statistics-theory and Methods 3: 1-27.

Collins, A. M., & Quillian, M. R. (1969). Retrieval Time from Semantic Memory. Journal of Verbal Learning and Verbal Behavior, 8, 240–247.

Danon, L., Díaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. Journal of Statistical Mechanics, 2005(9), P09008. https://doi.org/10.1088/1742-5468/2005/09/P09008

Granovetter, M. (1973). The strength of weak ties. American Journal of Sociology. Retrieved from http://www.jstor.org/stable/2776392

Granovetter, M. (1978). Threshold Models of Collective Behavior. American Journal of Sociology, 83(6), 1420–1443.

Houghton, J., Siegel, M., Madnick, S., Tounaka, N., Shirnen, B., Nakagawa, D., Nakamura, K., Sugiyama, T. (Forthcoming) Beyond Keywords: Tracking the evolution of conversational clusters in social media. Sociological Methods and Research.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’03, 137. https://doi.org/10.1145/956755.956769

Kossinets, G., & Watts, D. J. (2009). Origins of Homophily in an Evolving Social Network. American Journal of Sociology, 115(2), 405–450.

Larose, D. T., & Larose, C. D. (2014). Discovering Knowledge in Data: An Introduction to Data Mining. Journal of Chemical Information and Modeling (2nd ed.). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1017/CBO9781107415324.004

Linn, M. C., Davis, E. A., & Bell, P. (2004). Internet environments for science education. British Journal of Educational Technology (Vol. 36).

Mayer, R. E. (1995). The Search for Insight: Grappling with Gestalt Psychology's Unanswered Questions. In R. J. Sternberg & J. E. Davidson (Eds.), The Nature of Insight.

Mcpherson, M., Smith-lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 69(2 2), 1–15. https://doi.org/10.1103/PhysRevE.69.026113

Parthasarathy, S., Ruan, Y., & Satuluri, V. (2011). Community Discovery in Social Networks: Applications, Methods, and Emerging Trends. In C. C. Aggarwal (Ed.), Social Network Data Analytics (pp. 79–113). Springer Science. https://doi.org/10.1007/978-1-4419-8462-3

Posner, M. I., Petersen, S. E., Fox, P. T., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. Science (New York, N.Y.), 240(4859), 1627–1631. https://doi.org/10.1126/science.3289116

Schilling, M. A. (2005). A "Small-World" Network Model of Cognitive Insight. Creativity Research Journal, 17(2–3), 131–154. https://doi.org/10.1080/10400419.2005.9651475

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. Sociological Methods & Research, 40(2), 211–239. https://doi.org/10.1177/0049124111404820

Sowa, J. F. (1991). Principles of Semantic Networks: Explorations in the Representation of Knowledge. Morgan Kaufmann series in representation and reasoning.

Speer, R., & Havasi, C. (2012). Representing General Relational Knowledge in ConceptNet 5. Lrec, 3679–3686.

Travers, J., & Milgram, S. (1969). An Experimental Study of the Small World Problem. Sociometry, 32(4), 425–443. Retrieved from http://www.jstor.org/stable/2786545

Tulving, E. (1972). Episodic and semantic memory. Organization of Memory. https://doi.org/10.1017/S0140525X00047257

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of Embeddedness. Administrative Science Quarterly.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. Nature, 393(6684), 440–2. https://doi.org/10.1038/30918

Watts, D. J. (2002). A simple model of global cascades on random networks. Proceedings of the National

Academy of Sciences, 99(9), 5766–5771. https://doi.org/10.1073/pnas.082090499

Woods, W. A. (1975). WHAT'S IN A LINK: Foundations for Semantic Networks. Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence, (November), 102–125.

Yang, B., Cheung, W. K., & Liu, J. (2007). Community mining from signed social networks. IEEE Transactions on Knowledge and Data Engineering, 19(10), 1333–1348. https://doi.org/10.1109/TKDE.2007.1061

Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. Journal of Anthropological Research, 33(4), 452–473.

---

1. The parameters which govern Bob's process are chosen to maximize the clarity of this exposition. In subsequent analysis, each of the parameter values described here is explored to determine the sensitivity of conclusions to particular values. ↩
2. A test of the sensitivity of this paper's main conclusions to changes in this scoring system is described in Appendix 1. ↩
3. An alternative formulation could involve selecting the items to be excluded in proportional to their length score to the preference power. This would merely tend to weaken the effect of the preference. For the sake of simplicity the single formulation is used. ↩