Structural Diversity in Social Recommender Systems

Xinyi (Lisa) Huang
University of Waterloo
x37huang@uwaterloo.ca

Mitul Tiwari
LinkedIn
mtiwari@linkedin.com

Sam Shah
LinkedIn
samshah@linkedin.com

ABSTRACT

Online social networks have become important for sharing, discovery, communication, and networking. Recommender systems are an essential part of any social network. For example, recommending people to connect with is essential for the growth of the network since an online social network is only partially observed and two people might know each other but may not be connected. In this paper, we analyze data from LinkedIn, the largest online professional social network, which recommends other members to connect through its "People You May Know" feature. Analyzing the effect of structural diversity on the invitation rate from such member recommendations, we find that higher connection density and lower structural diversity results in a higher connection invitation rate. We also analyze and study the effects of structural diversity of members' connection networks on their engagement on the LinkedIn network.

General Terms: Social Recommender Systems, Structural Diversity, Engagement

1. INTRODUCTION

Social networks commonly have a way of expressing a link between members: LinkedIn uses bidirectional links called connections, Facebook has bidirectional links denoting friendships, while Twitter and YouTube have unidirectional links for followers and subscribers respectively. Connections are important in establishing relationships and forming communities in a network, which usually induces higher engagement in an individual member and virality of network activities.

This paper investigates how the connections in a social network can influence member activity and engagement. Specifically, we analyze how the *structural diversity*—the number of connected components in a set—can affect members' decision to send connection invitations and to engage in network activities.

Recent research using the Facebook social network indicates that increasing structural diversity improves user adoption and engagement [18]. The research hypothesizes that these measures could be applied to other networks and problem domains.

To that end, we investigate the influence of structural diversity in other contexts. LinkedIn's "People You May Know" feature attempts to find other members a user may know on the social network. This is a link prediction problem where node and edge features in the social graph are used to predict whether two people know each other. We compare the rate of invitations sent from LinkedIn members based on the structural diversity of a recommended set of potential connections. We also investigate invitations through member-uploaded contacts, which elides the bias of a recommendation model. Here, a member uploads her contact list and is then presented with a set of matching members. In addition, we also apply analysis of structural diversity of a members' network, comparing it against a member's engagement.

Our analysis shows that higher connection density and lower structural diversity in a recommended set results in a higher invitation rate, which presents a conflicting view to recent research. In this paper, using similar metrics for assessing structural diversity, but for slightly different use cases, we present contrasting trends for invitation rates, though reinforce trends for engagement rates. Our analysis shows that the effect of structural diversity in a recommender system is use case dependent and requires care to generalize in other recommendation contexts.

2. RELATED WORK

There is plenty of research on recommendation systems in a social network and consequently on quality improvements of recommendations [2, 3, 6, 7, 9–11, 13, 15, 16, 18, 19]. One of the most common techniques is comparing the similarity of user content and friends-of-friends for recommendations to social network members [5, 10]. Research suggests that friendship and community membership have a strong influence on recommender systems in a social network [21]. Specifically, individuals likely take action once a number of members in their immediate network is seen to have taken the same action [16]. The structure in community or group affiliations can thus be used to closely predict the structure of the respective person-to-person network and even their actions [17]. Moreover, recent research developments analyze how structural diversity can affect the quality of recommendations [1, 2, 5, 11, 14, 20, 22]. Member satisfaction and the effectiveness of recommendation sets not only depend on the accuracy of predictions, but also on the diversity of the set [22]. The lack of diversification, in some cases, can lead to over-specialization and does not provide sufficient coverage of the domain of recommendations [14, 20]. However, some experiments show that users have a higher acceptance rate when recommended to people from within the same company divisions [5]. Ugander et al. [18] suggest that having more diverse structure in recommendations leads to higher invitations and that having a more diverse friendship network leads to more user activity on Facebook. Here, diversity is measured based on the number of components in the social graph. Using similar metrics for assessing diversity, this paper presents

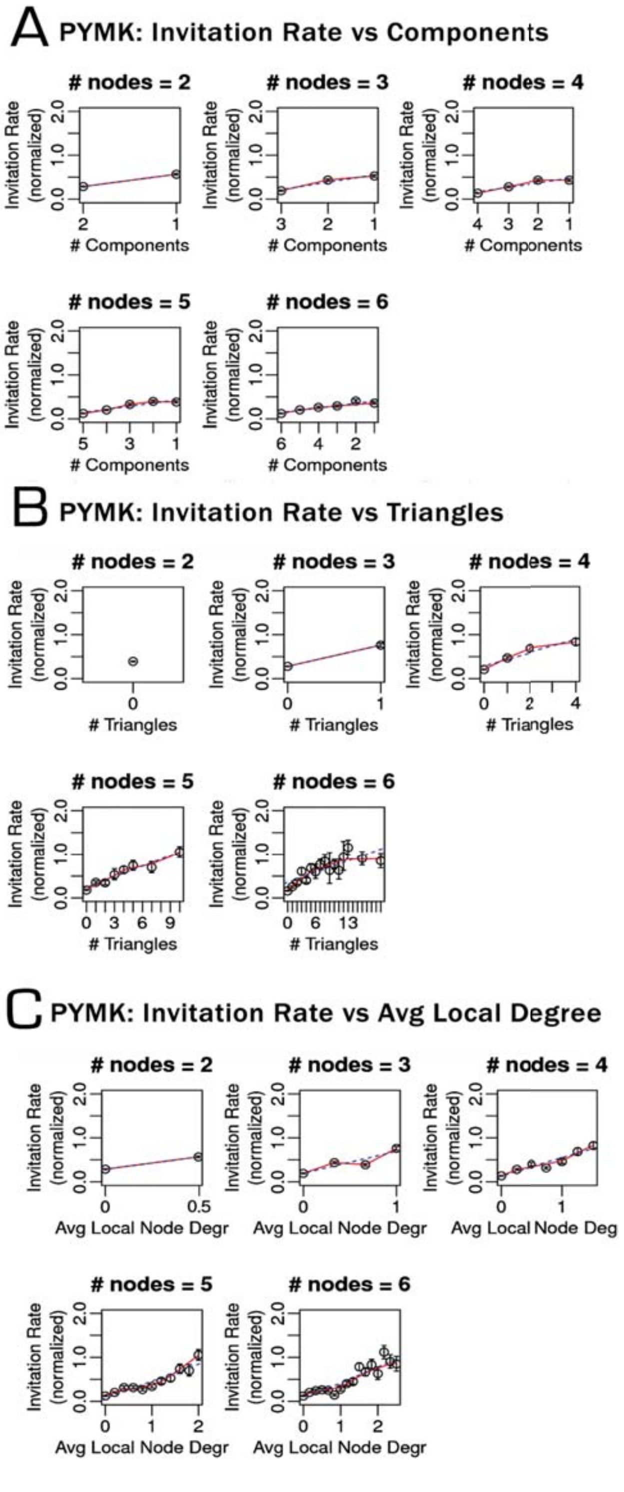

Figure 1: Invitation rates over a month of "People You May Know" (PYMK) recommendations of sizes 2-6. (A) Invitation rates compared to the number of connected components. (B) Invitation rates compared to the number of triangles. (C) Invitation rates compared to the average local node degree. Invitation rates are reported on a relative scale, with 1.0 representing the mean of the results.

a different trend for invitation rates but reinforces the trend for engagement rates. We cover slightly different recommendation use cases and show new aspects to this field of research.

3. BACKGROUND

Recommender systems are one of the tools that social networks employ to increase member activity and network connectivity. These systems work by analyzing current knowledge of a member's data and making assumptions about the data that has not yet been collected. Specifically, many recommender systems make use of the social network graph to make educated guesses as to whether two unconnected members might actually be acquainted. With sufficient data collected, the recommender system could provide members with connection suggestions to grow their network.

A social network maps to a graph where vertices represent people in the network, and an edge represents the connection between two people. It is common to use trusses, cliques, and components to determine the overall connectedness or density of a graph structure [2, 8]. Existing research has often used the number of trusses or cliques in defining the structural diversity of a network, where a lower number of such subgraphs implies greater diversity [4]. Structural diversity can be also defined in terms of the number of components in a graph: a higher number of components implies greater structural diversity [8, 18]. Furthermore, components are sometimes quantified through k*-core* or k*-brace* decompositions of the network graph to eliminate influence from unimportant nodes [12]. The k-core decomposition of a graph is its subgraph induced by repeatedly removing nodes with fewer than k neighbors [4, 12]. The k-brace of a graph is obtained by deleting all edges with *embeddedness* less than k, where the embeddedness of an edge is the number of common neighbors shared by its two endpoints [4].

4. RECOMMENDATIONS

Any online social network is partially observed; that is, two people might know each other but might not be connected with each on the site. To increase connectivity and form communities, it is common for social network websites to recommend members to connect with. Often, the implementation of such recommendation engines is designed to show recommendations that are most relevant and appealing to members. Research has shown that the structure within the recommendation sets also have an effect on an action [3, 18].

4.1 People You May Know

To encourage higher connectivity and activity between LinkedIn's members, recommendation features such as "People You May Know" (PYMK) suggest potential connections for members and encourage them to send invitations to connect. In this work, we analyze the relationship between the structural diversity of PYMK recommendations sets and the rate of invitations.

4.1.1 Data Set and Experimental Setup

Our data set consists of PYMK recommendations that were shown to LinkedIn members over the course of a month. We also focused on PYMK recommendation sets of sizes 2-6 to better compare our results with respect to Ugander et al. [18].

Each recommendation set consisting of members is mapped to a graph G with the vertices representing the individual members being recommended, and edges representing connectivity on LinkedIn. The factors that measure structural diversity of the recommendation set are the number of *connected components*, number of *triangles*, and the *average local node degree* found in graph G. A *connected component* is a maximal subgraph in which every pair of vertices is

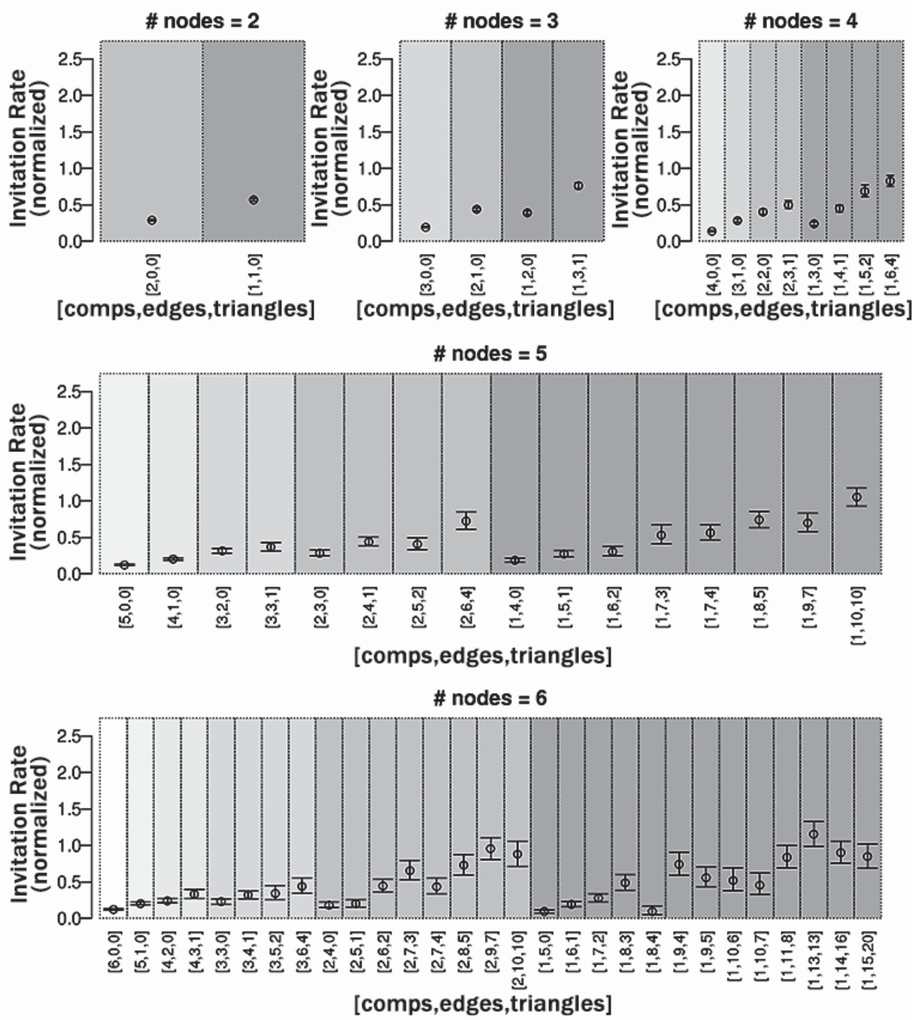

Figure 2: Invitation rates over a month of "People You May Know" (PYMK) recommendations of sizes 2–6 measured against an aggregate of components, triangles, and average local node degree. The measure on the x-axis represents density: the number of components, edges, and triangles were ordered in descending, ascending, and ascending orders, respectively, to align with increasing graph density. This aggregate gives a more detailed view of the graph structure beyond the individual measures. Invitation rates are reported on a relative scale, with 1.0 representing the mean of the results.

connected on a path (no isolated vertices) or is itself an isolated vertex. A *triangle* is a set of three vertices in the graph G in which each vertex is connected to the other two. Note that if graph G contains only two nodes then there are no triangles. The *local node degree* is the number of connections a node has in graph G, and the *average local node degree* takes an average of these numbers. The concept of local node degree is also referred as degree centrality in literature. High diversity corresponds to a large number of components, a low number of triangles, or a low average local node degree.

We also gathered connection invitation requests sent by LinkedIn members over the same period from the PYMK feature. We computed the *invitation rate*, the ratio of the number of connection invitations to the number of recommendations shown, to measure overall effectiveness.

4.1.2 Experimental Results

Our experiments compute invitation rate over this PYMK recommendation data set, and we compare the invitation rate to the three aforementioned factors that measure structural diversity. This is shown in Figure 1. The gathered data shows that more invitations are sent with recommendation sets with fewer connected components, a higher number of triangles, and a higher average local node degree. In the initial experiments, we considered all of this PYMK recommendation data set and corresponding invitation rate, with similar results to Figure 1. In this figure, we excluded cases where a member does not even look at the recommendation and leaves the page; that is, invitation rates were generated for instances where at least one invite was sent out.

We have also aggregated the three measures of structural diversity to establish a trend as shown in Figure 2. The same general trend can be observed from the aggregate measure: the less diverse the recommendation set, the higher the invitation rates tend to be. A possible reason is that if a member is acquainted with one person in a recommendation set of closely connected peers, they will likely also be acquainted with several others in the set due to high connection density. This consequently results in the propagation of

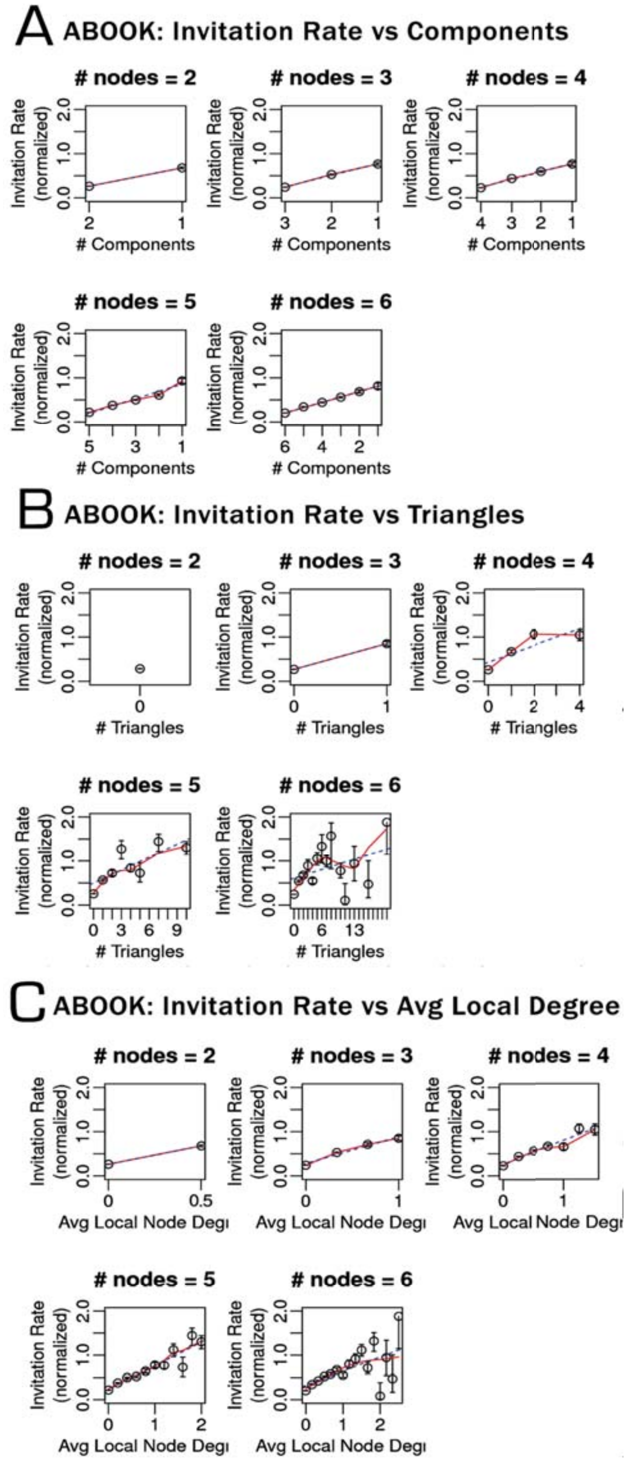

Figure 3: Invitation rates over a month for address book contacts imports (ABOOK) of sizes 2–6. (A) Invitation rates compared to the number of connected components. (B) Invitation rates compared to the number of triangles. (C) Invitation rates compared to the average local node degree. Invitation rates are reported on a relative scale, with 1.0 representing the mean of the results.

connection invitations, generating high invitation rates.

4.1.3 *Discussion*

Ugander et al. [18] suggests that a more diverse set of recommendations translate to higher sign-up rates based on Facebook's member recruitment data. Although our experiment produced different findings, it is evident that these are two different use cases: joining a social network based on recommendations versus connecting with others in the network through recommendations. In the former case, it can be argued that the greater the diversity, the more likely members will see a recommendation that would make them want to join Facebook. For instance, many individuals have closer relationships with their university or company networks than their high school networks. These individuals may then feel more enticed to join Facebook if shown diverse recommendations containing friends from their high school, university, and company than if shown a densely connected group of high school friends. On the contrary, densely connected recommendations on LinkedIn exhibit a propagation effect; one invitation is likely to lead to another. This is because knowing one person in a densely connected group usually implies knowing that person's close connections in the group as well. Based on this analysis, we can infer that the effect of strucural diversity in a recommender system highly depends on the corresponding use case.

4.2 Contacts Import

Because the data gathered from PYMK recommendations showed a different trend than the Facebook recruitment recommendations, we considered the possibility that PYMK bias might have affected invitation rates. We also explored contacts import invites, which unlike PYMK, has no bias because it simply takes contact information uploaded by members. Contacts import is a feature on LinkedIn that enables new members to upload their contact address books to find potential connections. With contacts import invites, we performed the same experiment by comparing the invitation rate to structural diversity of members uploaded in contacts. The results of these experiments are shown in Figures 3 and 4.

We found the same trend of higher invitation rates with higher network density. Therefore, we can conclude that based on LinkedIn's member data, recommendation sets with lower diversity can be associated with higher invitation rates.

5. ENGAGEMENT

Social networks promote connectivity for the subsequent effect of increasing user engagement. In addition to several interest and social factors, social engagement can be dependent on the structure of one's immediate connections network. We take a LinkedIn member, and form a graph G' where the nodes are composed of people in the member's network and edges represent if these people are connected. Consequently, metrics similar to that of the recommendation sets can be used to investigate correlations between the engagement of members and the structural diversity of their connections.

Analysis similar to that of recommendations was applied to user engagement using LinkedIn's member data. Engagement was measured based on the number of page views, where users are considered "engaged" if they have visited LinkedIn on at least some fixed number of days during a week. Ugander et al. [18] defined engagement similarly using some threshold in weekly Facebook visits to consider users as engaged. Both analyses were done on users with 10, 20, 30, 40, or 50 connections or friends in their social network. The results on LinkedIn's data is shown in Figure 5. Using the number of connected components as a measure of diversity for LinkedIn's user engagement, it appears that the trend converges on

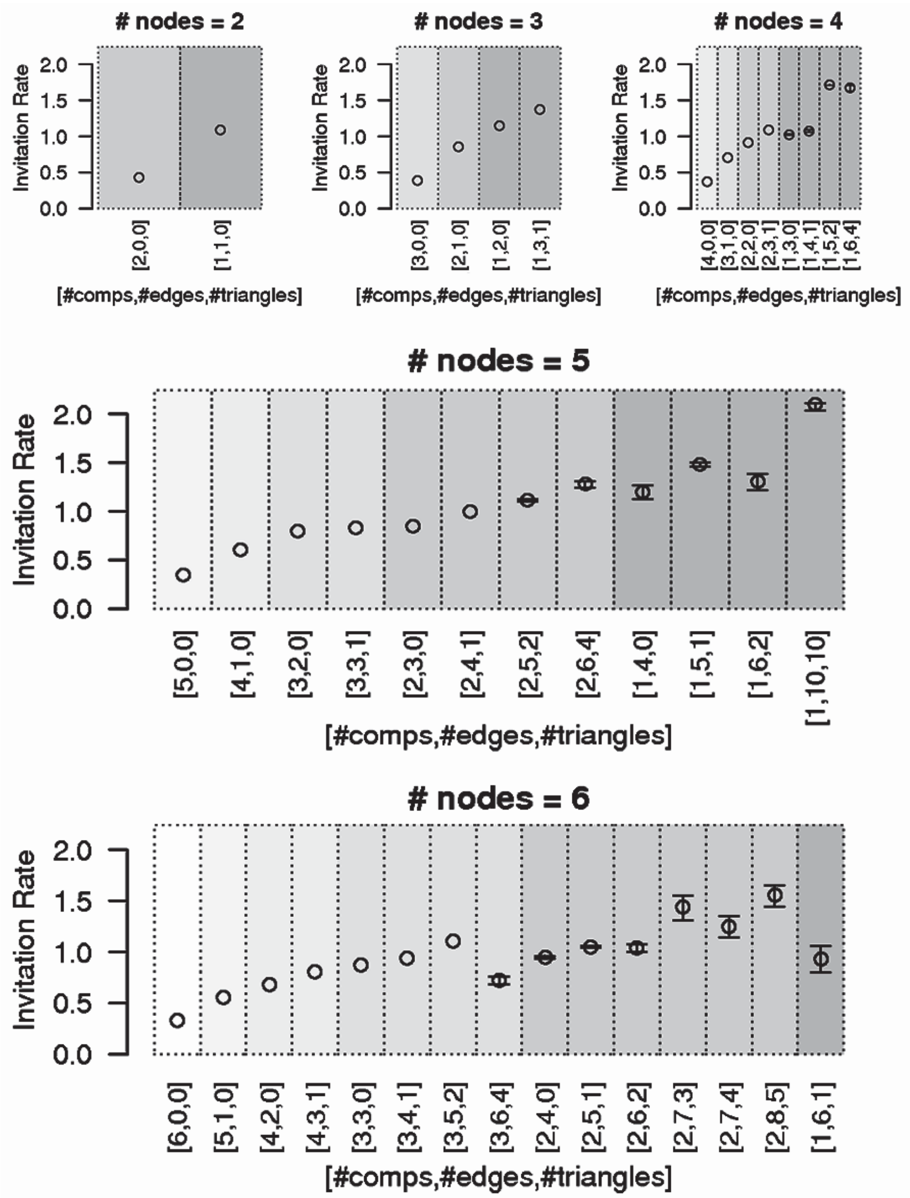

Figure 4: Invitation rates over a month of contact imports of sizes 2-6 measured against an aggregate of components, triangles, and average local node degree. The measure on the x-axis represents density: the number of components, edges, and triangles were ordered in descending, ascending, and ascending orders respectively, to align with increasing graph density. This aggregate gives a more detailed view of the graph structure beyond the individual measures. Invitation rates are reported on a relative scale, with 1.0 representing the mean of the results.

lower engagement rates as the number of components increases. This differs from the findings for LinkedIn recommendations but agrees with the Facebook finding. Similar to Facebook engagement analysis, we dismissed nodes of lower degree (assumed to have lower importance and relevance) by applying k-core decomposition to the connections graph and found that a higher number of k-core components (experimented with k=1 and k=2) translate to higher engagement. These results coincide with what Ugander et al. [18] conclude through using k-core decompositions to analyze engagement trends.Therefore, our research reinforces the theory that more diverse connections networks result in higher user engagement.

6. CONCLUSION

In this paper, we investigate how the structural diversity of connections in a social network can affect members when deciding to send connection invitations and to engage in network activities. We compared the rate of invitations sent from LinkedIn members based on the structural diversity of a recommended set of potential connections. Our analysis shows that higher connection density and lower structural diversity in a recommended set results in a higher connection invitation rate, which presents a contrasting trend compared to a recent study using Facebook data for a different use case, recruitment emails [18]. Our investigation into the effect

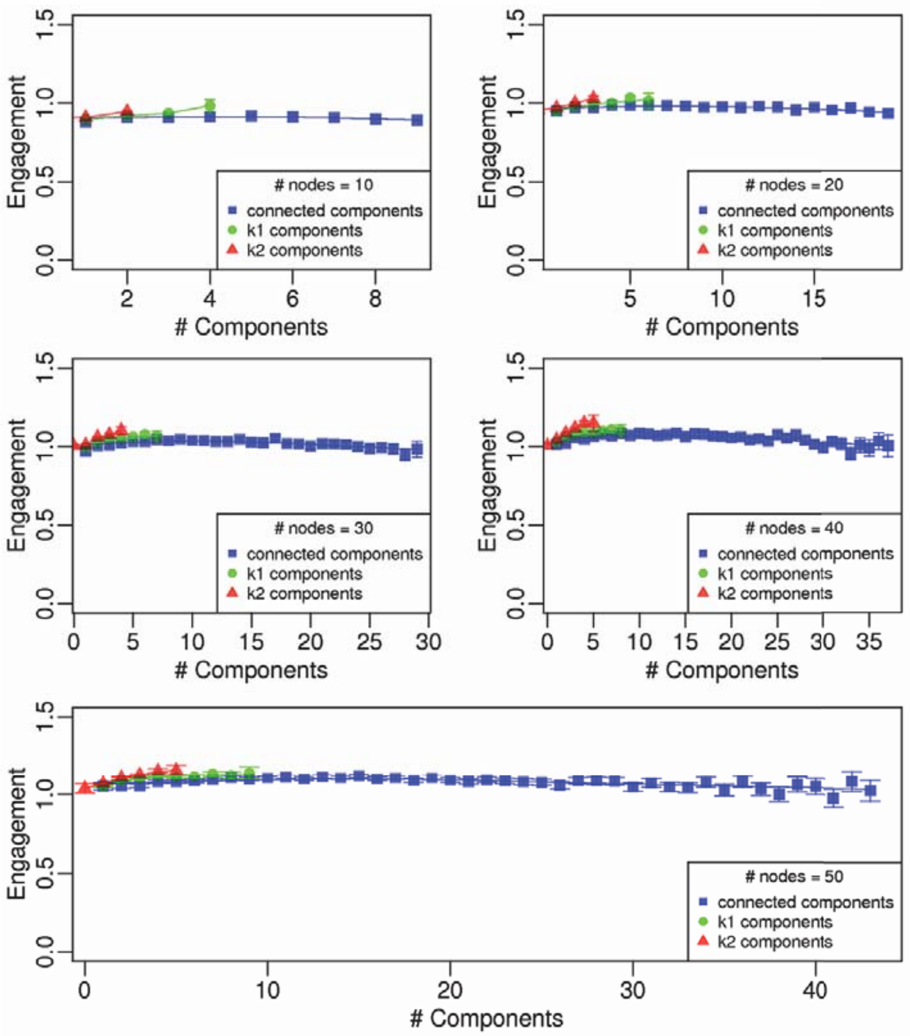

Figure 5: Engagement rates measured against connected components, k1 components, and k2 components. Data was collected over a 6-week period for members with a network of sizes 10, 20, 30, 40, and 50. Engagement rates are reported on a relative scale, with 1.0 representing the mean of the results.

of structural diversity of a member's connection network on a member's engagement found that higher structural diversity results in higher engagement, which is similar to previous findings. We conclude from our analysis that the effect of structural diversity in a recommender system highly depends on the corresponding use case and it would be a mistake to generalize the effects of structural diversity on one use case of recommender system to all use cases.

References

[1] Gediminas Adomavicius and YoungOk Kwon. Toward more diverse recommendations: item re-ranking methods for recommender systems. In *Proceedings of the 19th Workshop on Information Technologies and Systems*, 2009.

[2] Gediminas Adomavicius and YoungOk Kwon. Maximizing aggregate recommendation diversity: A graph-theoretic approach. In *Proceedings of the 5th ACM conference on Recommender Systems*, 2011.

[3] Lars Backstrom. Group formation in large social networks: membership, growth, and evolution. In *KDD*, 2006.

[4] Jonathan Cohen. Trusses: Cohesive subgraphs for social network analysis. In *Proceedings of the National Security Agency Technical Report*, 2008.

[5] Elizabeth M. Daly, Werner Geyer, and David R. Millen. The network effects of recommending social connections. In *Proceedings of the 4th ACM Conference on Recommender Systems*, 2010.

[6] Peter Sheridan Dodds and Duncan J. Watts. Universal behavior in a generalized model of contagion. In *Proceedings of the Phys. Rev. Lett.*, 2004.

[7] Ilham Esslimani, Armelle Brun, and Anne Boyer. Densifying a behavioral recommender system by social networks link prediction methods. *Social Networks Analysis and Mining Journal*, 1(3):(159–172), 2011.

[8] Linton C. Freeman. Cliques, galois lattices, and the structure of human social groups. *Social Networks*, 18:173–187, 1996.

[9] Jennifer Golbeck and Derek L. Hansen. A framework for recommending collections. In *Proceedings of the 5th ACM conference on Recommender Systems*, 2011.

[10] Ido Guy, Inbal Ronen, and Eric Wilcox. Do you know?: recommending people to invite into your social network. In *Proceedings of the 14th international conference on Intelligent user interfaces*, 2009.

[11] Rong Hu and Pearl Pu. Helping users perceive recommendation diversity. In *Proceedings of the ACM TIST Special Issue on Diversity and Discovery in Recommender Systems and Exploratory Search (DiveRS)*, 2011.

[12] Svante Janson and Malwina J. Luczak. A simple solution to the k-core problem. *Random Struct Algo*, 30:50–62, 2007.

[13] David Krackhardt. Cognitive social structures. *Social Networks*, 9:109–134, 1987.

[14] Lorraine Mcginty and Barry Smyth. On the role of diversity in conversational recommender systems. In *Proceedings of the Fifth International Conference on Case-Based Reasoning*, pages 276–290. Springer, 2003.

[15] James Moody and Douglas R. White. Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review*, 68(1):103–127, 2003.

[16] Elchanan Mossel and Sebastien Roch. On the submodularity of influence in social networks. In *Proceedings of the First Asian Pacific Conference on Intelligent Agent Technology, World Scientific Publishing*, 2007.

[17] Mark E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E.*, 64(2):026118 (1–17), 2001.

[18] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences*, 109(16): 5962–5966, 2012.

[19] Duncan J. Watts and Peter Sheridan Dodd. Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34:441–458, 2007.

[20] Cong Yu, Laks Lakshmanan, and Sihem Amer-Yahia. It takes variety to make a world: diversification in recommender systems. In *Proceedings of the International Conference on Extending Database Technology*, 2009.

[21] Quan Yuan, Li Chen, and Shiwan Zhao. Factorization vs. regularization: fusing heterogeneous social relationships in top-n recommendation. In *Proceedings of the 5th ACM conference on Recommender Systems*, 2011.

[22] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, 2005.