Triad-based Role Discovery for Large Social Systems

Derek Doran

Department of Computer Science & Engineering
Kno.e.sis Research Center
Wright State University, Dayton, OH
derek.doran@wright.edu

Abstract. The *social role* of a participant in a social system conceptualizes the circumstances under which she chooses to interact with others, making their discovery and analysis important for theoretical and practical purposes. In this paper, we propose a methodology to detect such roles by utilizing the conditional triad censuses of ego-networks. These censuses are a promising tool for social role extraction because they capture the degree to which basic social forces push upon a user to interact with others in a system. Clusters of triad censuses, inferred from network samples that preserve local structural properties, define the social roles. The approach is demonstrated on two large online interaction networks.

1 Introduction and Motivation

Why do users choose to participate and interact with others in a social system? This fundamental question lies at the heart of many sociological studies that examine the way people interact within a community. A *social role* is a powerful conceptualization for reasoning about the nature of these interactions and can be used to infer why users choose to participate. It is defined as a descriptive label that expresses the circumstances and reasons under which a user interacts with others in a community [1]. Social roles determine the set of interaction partners of an ego, and have a direct affect on choose they interact with. The concept is theoretically based on a notion of the user's *position* in a network of interactions; how one decides to embed themselves among others, based on who they choose to forge relationships with, can explain how they are perceived and their ability to spread information or influence [2]. Such perceptions and abilities factor into why and how a user interacts with others [3]. Practically, the delineation of users by their social role facilitates the analysis and interpretation of complex social networks by simplifying them from interactions among users to interactions among roles [4]. It also lets researchers perform comparative studies of different communities by comparing the structure of interactions among roles common to many contexts. Role analysis can also help us identify the kinds of roles (and hence users) that may become influential [5], and reveal latent social structures within social systems [6]. Furthermore, meta-analysis of the kinds of roles and the interactions among them can help designers create effective physical and digital spaces for communities and organizations to grow within [7].

Two users exhibit the same social role if they are in "regularly equivalent" positions [8]. Finding such positions, however, is analogous to searching for a k-coloring of a network with k unknown a priori (an NP-hard problem [9]). The vast scale of many online social systems thus make it infeasible to precisely identify the social roles within them. Researchers have instead turned to approximation methods that define roles based on the *structural similarity* of users' ego-networks [7, 10, 11]. Such methods capture the notion that users exhibiting *similar patterns of interactions with others* contribute towards and utilizes a social system in comparable ways, and thus, take on similar social roles. However, present approaches find similarities among structural ego-network features that reflect their overall shape, instead of micro-level features that better reflect a users' embedding within their peers. The resulting groups of social roles may thus consist of discordant ego-networks with few common interaction patterns and motifs. Some methods even need to apply further, potentially distorting post-processing steps [11, 10] to the roles that are mined.

This paper introduces a new approach to detect the social roles users exhibit in large social systems. It evaluates the similarity of ego-networks according to their *conditional triad census*, which is a vector capturing the different types of three way relationships it is composed of. This representation holds more promise for discovering roles becacuse triad types are indicative of specific sociological forces that drive interactions at a basic level, and hence, speak closely to the social role concept. Users are ground into roles by clustering the conditional triad censuses of their ego-networks. Two large social systems are used to test our approach: an online social network and collaborative editing platform.

This paper is organized as follows: Section 2 gives an overview and assessment of the related work. Section 3 introduces the concept of a conditional triad census. Section 4 presents a method to detect roles based on conditional triad censuses. Section 5 analyzes the structure of the social roles mined from large social systems. Conclusions and directions for future work are offered in Section 6.

2 Related Research

Broadly, previous work for studying social roles in large or online social systems may be divided into two types: (i) implied role analysis; and (ii) automatic social role extraction. Implied role analysis predefines the set of social roles users in a social system are expected to exhibit based on an analyst's understanding of how interactions within it occur. For example, Nolker *et al.* predefine members of a Usenet group into the roles *leader, motivator*, and *chatter* [12] based on their own hypothesis about the nature of Usenet interactions. Golder *et al.* also studied Usenet groups but proposed a different taxonomy of roles that include *celebrity, ranter, lurker* and *troll* [13]. Gliwa *et al.* examined collections of online bloggers and defined roles such as *selfish influential user, social influential blogger,* and *standard commentator* [5]. Welser *et al.* defined the roles *substantive experts, technical editors, counter vandalism,* and *social networkers* for Wikipedia users [14]. These implied role analyses are based off of social roles that are presumed to exist without evaluating any interactions in the network first. Thus, studies may define different sets of implied roles over the same kind of online social system, inducing conflict or confusion. For example, it is unclear if the Usenet roles *leader, motivator*, and *chatter* [12] compatible with the alternative set *celebrities, ranters, lurkers, trolls,* and *newbie* [13], when one set is more suitable than the other, and if the cross-product of the two sets (e.g. *leader-celeberty; chatter-lurker,* etc.) is a valid collection of roles. Furthermore, implied role analyses search for evidence of the roles they assert to exist prior to analysis. However, one can find statistically significant evidence for almost any model when data sets are very large [15].

Instead, automatic role extraction methods lets a social system 'speak for itself' by defining roles purely based on observed data. Hautz *et al.* categorized users in an online community of jewelry designers by mapping out- and in-degree distributions and frequency of interactions to "low" and "high" levels [7]. Zhu *et al.* identify social roles across phone call networks based on ego-network clustering coefficients and mean geodesic distances between users [10]. Chan *et al.* discover roles using over fifty behavioral and structural features of users across the post/reply network of online forums [11]. However, such role extraction methods commonly use quantitative structural and behavioral features that may not speak to the nature of a user's role. For example, clustering coefficients and degree distributions quantify the totality of an ego-network's structure, even though it is the specific kinds of interaction patterns within them that reflect one's social role [3]. To overcome this limitation, we next introduce a role extraction method that represents of ego-networks by their conditional triad census.

3 Conditional Triad Census

In social network analysis, a *triad* is defined as a group of three individuals and the pairwise interactions among them [16, 17]. They are the smallest sociological unit from which the dynamics of a multi-person relationship can be observed [18]. For example, third actors may act as a moderating force that modifies the relationship among two others [19]. Figure 1 captures the 36 different *conditional triads*, or ways an individual can be oriented towards two alters within a triad [20]. We define the *conditional triad census* of an ego-network is defined by a 36-element vector whose i^{th} component corresponds to the proportion of triads in the ego-network that are of type i.

Observational data has been used to develop theories that associate the configuration of a triad to underlying effects that promote specific kinds of social interactions [21, 22]. For example, triad 5 has an ego that receives interactions from two alters but chooses not to reciprocate. Ego-networks largely composed of this triad suggest that the ego receives many interactions but, for possibly selfish reasons, seldom chooses to reciprocate. By summarizing how frequently each kind of triad appears, a conditional triad census can thus model the strength of the different social forces that explain the nature of an ego's interactions. These forces, taken together, represent the circumstances and reasons why an ego participates in a social system, and thus, can explain her social role. Thus, we argue that searching for ego-networks whose conditional triad censuses are similar will lead to a meaningful grouping of users into social roles.

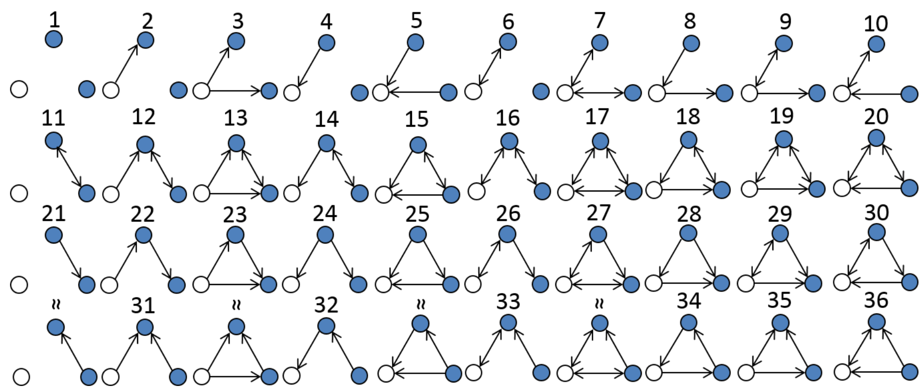


Fig. 1. Types of conditional triads

4 Extracting Social Roles

This section explains the process for extracting social roles based on conditional triad censuses. It first requires a careful sampling of the networks to make the computation of censuses scalable to the size of large social systems. k-means clustering is subsequently used to separate users by the social roles they exhibit.

We first introduce data from two popular online social systems, namely Facebook and Wikipedia, to demonstrate our methodology. These systems were chosen because users participate within them for different purposes, because of the distinct way interactions in the networks are defined, and because the interactions captured in our data represent strong associations. Facebook is used as a platform to informally share personal information, photos, and events with friends and family. We built its interaction network by placing a directed edge from user a to b if a posts at least one message on the wall (a collection of public messages) of b. Wikipedia is an online encyclopedia with articles that are written and edited by an open community of users. Interactions on Wikipedia are defined by the modification of content contributed by another user; we add a directed edge from a to b if a edited the text, reverted a change, or voted on approving an action to an article made by b. We built interaction networks for Facebook and Wikipedia using publicly available datasets [23, 24].

Table 1 presents summary statistics for these interaction networks, illustrating how they vary in size, shape, and user behaviors. The Facebook network has the smallest number of users (46,952). The Wikipedia network is almost three

Network	$\|V\|$	$\|E\|$	$\bar{d}$	α_{in}	α_{out}	$\bar{C}$
Facebook	46,952	264,004	37.36	1.61 ($p > 0.732$)	1.68 ($p > 0.964$)	0.085
Wikipedia	138,592	740,397	10.68	1.54 ($p > 0.999$)	1.83 ($p > 0.999$)	0.038

Table 1. Dataset summary statistics

times the size of Facebook, with 138,592 users and 740,397 distinct pairwise interactions, but its clustering coefficient is approximately 55% smaller. These measurements suggest that Facebook users have a greater tendency to surround themselves in within more connected ego-networks compared to Wikipedia users. The clustering coefficient of the Wikipedia is over half the size of the Facebook network. This could be explained by users who generally limit themselves to modifying articles written by a specific group (perhaps representing a specific topic), but could also be making minor edits (e.g., spelling or grammar) across the entire site. Users may thus have a tendency to organize into clusters based on their expertise, but because they also interact with all types of other groups by making simple technical edits, the clustering coefficient of the network is suppressed.

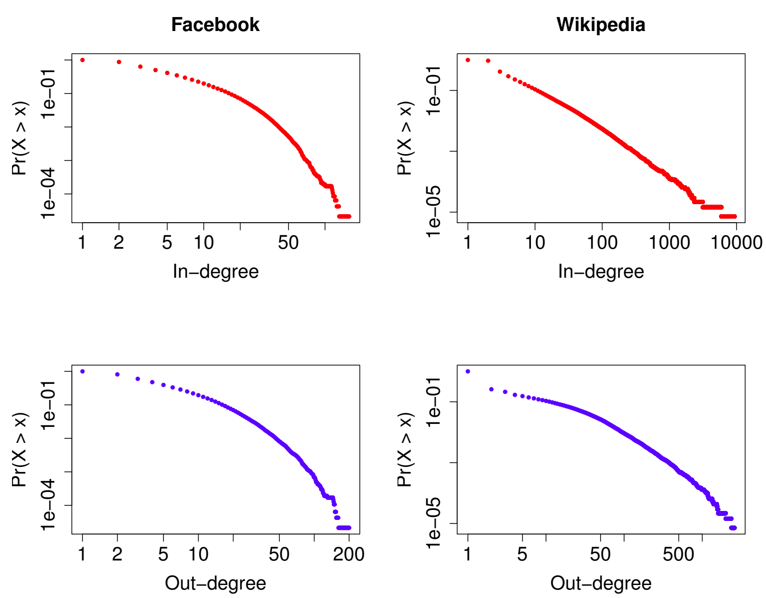
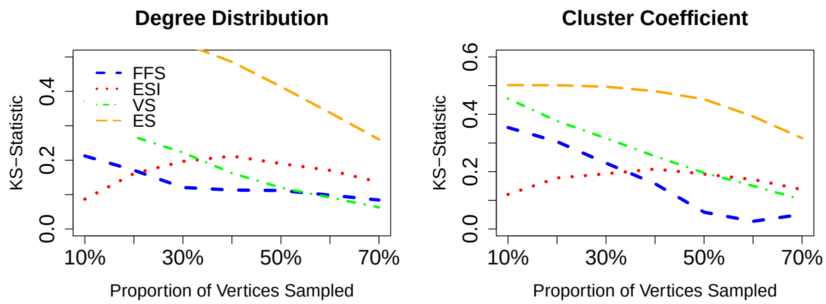
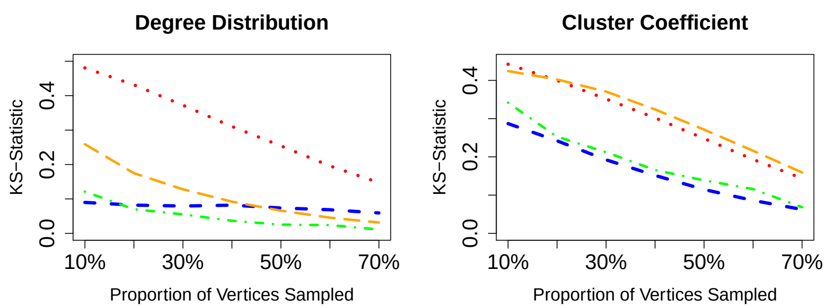


Fig. 2. Facebook (left) and Wikipedia (right) degree distributions on log-log scale

We also examine the in- and out-degree distributions of each network, presented in Figure 2, and note that they all exhibit a power-tailed shape. We test for power-law behavior using a maximum likelihood estimate approach [25] and list the resulting power-law exponent $\alpha_{in,out}$ in Table 1. We find the estimates of the power-law exponent to be very reliable ($p > 0.95$) except for the in-degree distribution of Facebook, which may be because its range only covers two orders of magnitude. All of the distributions exhibit a similar exponent ($1.54 < \alpha < 1.83$). A larger power-law exponent indicates that the distribution drops to zero faster in its right-tail [26], hence we are it is less likely to find a user who interacts with (receives interactions from) an unexpectedly high number of others on Wikipedia (Facebook) compared to Facebook (Wikipedia).



(a) Facebook



(b) Wikipedia

Fig. 3. Comparison of graph sampling methods

Network Sampling Computing the conditional triad census of every ego-network requires us to examine the configuration of $O(\binom{|V|}{3})$ triples in the network. Given the large size of these online social systems, even advanced algorithms that compute censuses in $O(|V|^2)$ [27] or $O(|E|)$ [28] time may be impracticably slow. However, since the components of a conditional triad census is the *proportion* of a triad type in an ego-network, the census of a smaller but similarly structured ego-network would yield a similar census. We therefore explore ways to approximate conditional triad censuses in the full network from a carefully selected network sample. A sample of a network G is a new network $G_s = (V_s, E_s)$ where $V_s \subset V$, $E_s \subset E$, and $|V_s| = \phi|V|$ with $0 < \phi < 1$. The configuration of triads within an ego-network critically rely on two local structural properties, namely, its degree distribution and the users' clustering coefficient. For example, ego-networks with high degree will naturally tend to have triads with relations among multiple alters (e.g. triads in rows 3, 5, and 7-10 in Figure 1), and lower (higher) cluster coefficients indicate a greater proportion of open (closed) triads.

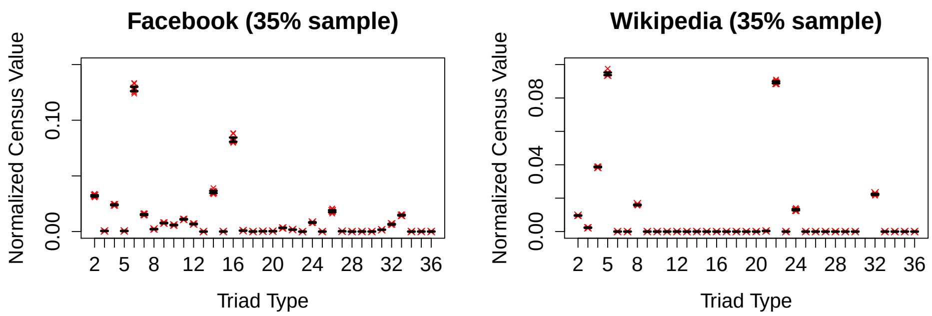


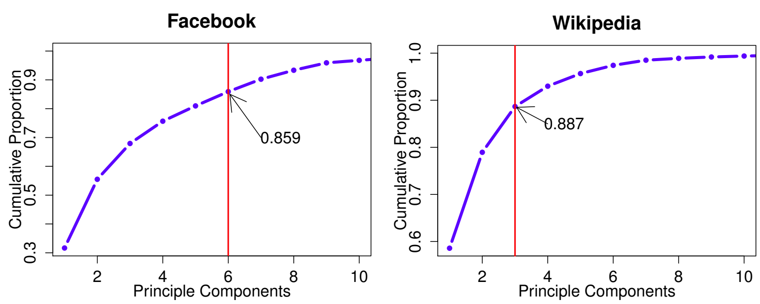
Fig. 4. Triad Role Census Sample Values with 95% Confidence Intervals

We consider four commonly used [29] graph sampling techniques for choosing V_s and E_s, and compare how well they are able to preserve the degree distribution of a users' ego-network and her clustering coefficient. These techniques are:

1. **Vertex Sampling (VS)**: Let V_s be a random sample of $\phi|V|$ vertices from V and define E_s to be the set of all edges among the vertices in V_s from G.
2. **Edge Sampling (ES)**: Randomly choose an edge $e = (v_1, v_2)$ from E, add it to E_s, and add v_1 and v_2 to V_s if they have not yet been added. Continue to choose edges from E until $|V_s| = \phi|V|$.
3. **Forest Fire Sampling (FFS)** [30]: Choose a random vertex v from V, randomly select $p/(1-p)$ of its outgoing edges, and add theses edges to E_s. Place every vertex incident to those added to E_s into a 'burned' set V_* and update V_s by $V_s = V_s \cup V_*$. Randomly choose a burned vertex from V_*, randomly select $p/(1-p)$ of its outgoing edges, and recursively repeat until $|V_s| = \phi|V|$. We use $p = 0.7$ based on the method author's suggestion [30].
4. **ES-i (ESI)** [29]: Randomly choose an edge $e = (v_1, v_2)$ from E and add v_1 and v_2 to V_s if they have not yet been added (note that e is not added to E_s). Continue sampling until $|V_s| = \phi|V|$. Finally, define E_s to be the set of all edges among the vertices in V_s from G.

We used the Kolmogorov-Smirnov distance metric $D = \sup_x |F_s(x) - F(x)|$ to compare how closely the degree and clustering coefficient distributions F_s of a sample taken with each method follow the distribution in the original network F. Figure 3 compares the average D of 100 samples taken at different levels of ϕ. We find that FFS does the best job at preserving both degree and clustering coefficient distributions for $\phi \geq 0.33$ on Facebook. It also best preserves the cluster coefficient distribution on Wikipedia, and performs similarly to VS in maintaining the degree distribution at $\phi \leq 0.35$. We further plot the average value of all conditional triad census components from $n = 20$ independently generated FFS samples of each network for $\phi = 0.35$ (we exclude triad 1 due to disproportionately high frequency) and their 95% confidence intervals in Figure 4. We find the sampling distribution of census proportions using FFS to have very small confidence intervals, indicating that the conditional triad censuses from any $\phi = 0.35$ sample is stable and may reasonably approximate the true censuses of the full network. We thus consider an FFS sample at $\phi = 0.35$ for the clustering analysis.

Census clustering We use k-means clustering, a common and flexible algorithm for discovering latent groups in data [31, 32], to separate users into roles. We use the ℓ^2-norm of the difference of two censuses from the FFS sampled network to measure their similarity. Since Figure 4 shows how many components are close or equal to 0, and hence are not useful dimensions to differentiate censuses, we reduce the dimensionality of our data using PCA [33]. As the scree plots in Figure 5 (a) and (b) show, we can reduce the complexity of the data from 36 to 6 and 3 dimensions for Facebook and Wikipedia, respectively, while preserving over 85% of the variation in the original data.



(a) (b)

Fig. 5. Principle Component Analysis Scree Plots

k-means clustering requires us to choose the number of clusters k to divide the data into beforehand. We use the silhouette coefficient metric [34] $SC_{\hat{C}^k}$ to evaluate differences in quality between divisions of censuses into k clusters. Larger values of $SC_{\hat{C}^k}$ correspond to a superior partitioning where the distance between clusters is large and the distance between vectors within cluster is small. For a given value of k, we ran 50 k-means clusterings over the PCA-reduced data using different random initializations of the centroid positions. Figures 6 (a) and (b) plots the average $SC_{\hat{C}^k}$ of these 50 trials for $2 \leq k \leq 9$. They reveal peaks at $k = 3$ and $k = 2$ clusters of the Facebook and Wikipedia censuses, respectively.

5 Role Analysis

To study the kinds of social roles that emerge from our clustering analysis, we identified the centroid positions C_i^* of each cluster C_i and searched for the user

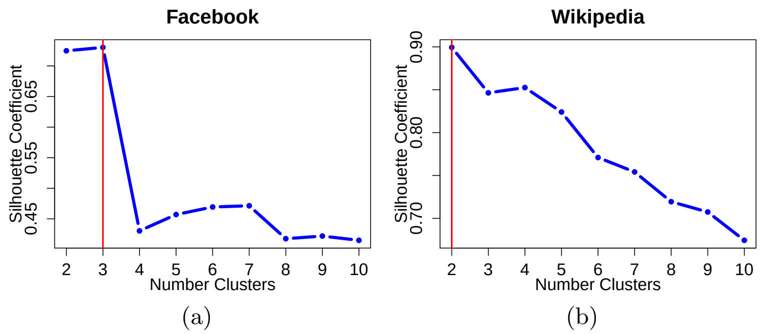


Fig. 6. k-means Clustering Silhouette Coefficients (bottom)

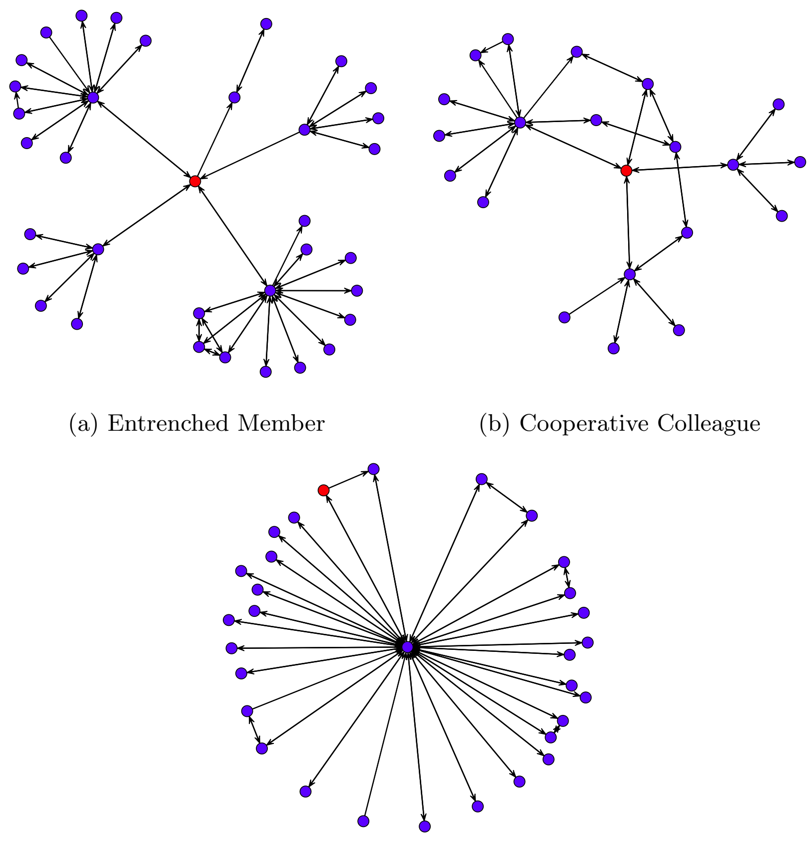
u_i^* whose conditional triad census is located closest to C_i^*. We define u_i^* as the "central user" of role i whose ego-network is the role's "central structure". Due to its position in the cluster, this "central structure" represents the way a prototypical user having this role embeds herself within the online social system.

Figure 7 presents the central role structures of the three social roles found on Facebook within the FFS $\phi = 0.35$ network sample. We label these roles and list the proportion of users falling under each role in Table 2. In these figures, the red node corresponds to the central user and the blue nodes are the members of their ego-network. Figure 7(a) represents a social role the majority of all Facebook users (56.6%) fall into with a central user who is embedded between many disconnected groups of others. They lie in an entrenched position critical for maintaining connectivity between the groups, and hence, act as a gatekeeper who can control the flow information from one group to another. However, given the fact that Facebook is used as a platform for social sharing, users may be embedded in such positions simply because it allows them to manage interactions within distinct groups of friends and contacts. For example, one can envision the user in Figure 7(a) to be sitting between groups that may correspond with colleagues at work, relatives, personal friends, and acquaintances. This structure may thus represent users that use the network to manage and facilitate communication with many non-overlapping social circles. By contrast, the 28.4% of users falling into the role represented by the central structure of Figure 7(b) find themselves surrounded by a web of interactions that occur between their first-degree connections. This minority of users cooperate with a variety of other interlinked colleagues participating in a free exchanged of information compared to the siloed, disconnected communities seen in Figure 7(a). participate in a single, tight-knit community of others, and thus correspond to users who use Facebook only to interact within a small group rather than as a tool for managing disconnected social circles.

Figure 7(c) corresponds to the 15% of users who are not embedded within a cohesive community or are entrenched between groups of others, but are positioned at the periphery of an active alter's neighborhood. These users thus ex-

Table 2. Mined social roles

Facebook	Role label	Structure	Proportion of users
	Entrenched Member	Figure 7(a)	56.6%
	Cooperative Colleague	Figure 7(b)	28.4%
	Casual Participant	Figure 7(c)	15.0%
Wikipedia	Role label	Structure	Proportion of users
	Specialist Attractor	Figure 7(d)	89.7%
	Generalist Attractor	Figure 7(e)	10.3%



(a) Entrenched Member

(b) Cooperative Colleague

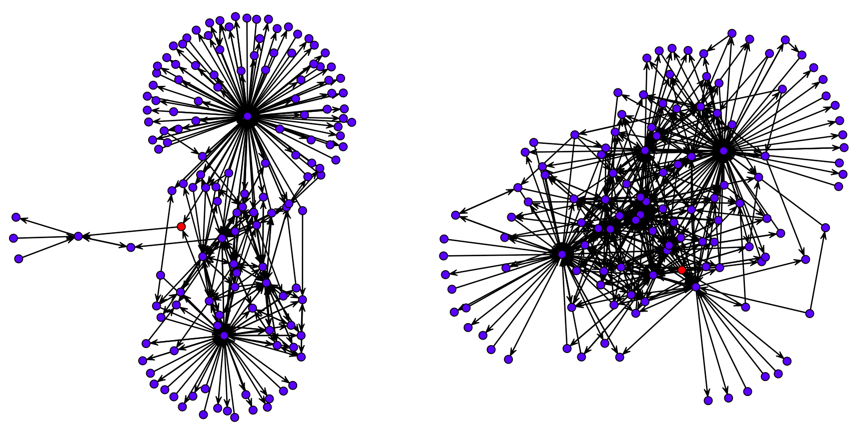
(c) Causal Participant

Fig. 7. Mined central role structures: Facebook

hibit little activity on the site, and are connected to one who prolifically shares information. This central structure thus suggests that these users are in a role that is not productive to the growth or dissemination of information; they are on

Facebook only to consume information from a single popular other, and hence simply participate in the network casually.

The density of the central structures of the two Wikipedia social roles shown in the bottom of Figure 8 is a result of the many different ways interactions are defined, as explained in Section 4. The social role taken on by the majority of users (89.7%) are represented by the structure in Figure 8(a). Here, we find the central user to be serving as a bridge between two sets of very active others (the two hubs) that are not strongly connected. It is interesting that these two others frequently edit articles written by many others, yet they seldom edit content added by the same individual. Such behaviors may emerge when two domain-specific specialists are only editing articles that only fall under their purview. This hypothesis is substantiated by previous research that found users labeled as substantive experts on Wikipedia exhibiting a similar ego-network structure [14]. Yet we still see a small amount of overlap between these two contributors, which may be done when editing Wikipedia articles that discuss many different topics. Since most Wikipedia articles do cover topics from multiple domains [35], we conclude that users falling under this role are contributors who write interdisciplinary articles that become edited by others with different expertise.



(a) Specialist Attractor (b) Generalist Attractor

Fig. 8. Mined central role structures: Wikipedia

The remaining 10.3% of users fall under the role whose structure represented in Figure 8(b). These users are positioned in the center of a highly connected structure, where almost every contribution is edited by a large number of others. We still observe users in hub positions that make edits to the work of separate collections of others, but unlike Figure 8(a), the hubs are strongly connected to each other. A plausible explanation for such structures are generalist editors who perform basic tasks such as spelling, grammar, and hyperlink corrections to contributed content that may have also been reviewed by specialists. Such technical editors may even be edited by other technical editors, as the language and wording of an article becomes more defined, adding to the density of this structure This explanation is also compatible with past observations of Wikipedia editors [14].

6 Conclusions & Future Work

This paper presented a novel methodology to detect social roles through the use of conditional triad censuses. This data-drive detection approach was applied to two different online social systems and extracted central structures that reflect intuitive reasons why users participate on Facebook and Wikipedia. This analysis made no assumptions about the roles users exhibit in the network prior to analysis. Future work will compare roles extracted by conditional triads over the same networks but under alternative definitions of network interactions. It will also analyze the differences between cluster-based role detection and using quantitative approximations of regular equivalence [36]. The feasibility of multiple role assignments with fuzzy clustering techniques will also be explored.

References

1. Lorrain, F., White, H.C.: Structural equivalence of individuals in social networks. The Journal of mathematical sociology **1** (1971) 49–80
2. Zaheer, A., Bell, G.G.: Benefiting from network position: firm capabilities, structural holes, and performance. Strategic management journal **26** (2005) 809–825
3. Scott, J., Carrington, P.J.: The SAGE handbook of social network analysis. SAGE publications (2011)
4. White, H.C., Boorman, S.A., Breiger, R.L.: Social structure from multiple networks. i. blockmodels of roles and positions. American journal of sociology (1976) 730–780
5. Gliwa, B., Zygmunt, A., Koźlak, J.: Analysis of roles and groups in blogosphere. In: Proc. of the 8th Intl. Conference on Computer Recognition Systems. (2013) 299–308
6. Laniado, D., Tasso, R., Volkovich, Y., Kaltenbrunner, A.: When the wikipedians talk: Network and tree structure of wikipedia discussion pages. In: Intl. Conference on Weblogs and Social Media. (2011)
7. Hautz, J., Hutter, K., Fuller, J., Matzler, K., Rieger, M.: How to establish an online innovation community? the role of users and their innovative content. In: Hawaii Intl. Conference on System Sciences. (2010) 1–11
8. Luczkovich, J.J., Borgatti, S.P., Johnson, J.C., Everett, M.G.: Defining and measuring trophic role similarity in food webs using regular equivalence. Journal of Theoretical Biology **220** (2003) 303–321
9. Khot, S.: Improved inapproximability results for maxclique, chromatic number and approximate graph coloring. In: Proc. of IEEE Syomposium on Foundations of Computer Science. (2001) 600–609

10. Zhu, T., Wang, B., Wu, B., Zhu, C.: Role defining using behavior-based clustering in telecommunication network. Expert Systems with Applications **38** (2011) 3902–3908
11. Chan, J., Hayes, C., Daly, E.M.: Decomposing discussion forums and boards using user roles. In: Intl. Conference on Weblogs and Social Media. Volume 10. (2010) 215–218
12. Nolker, R.D., Zhou, L.: Social computing and weighting to identify member roles in online communities. In: Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on. (2005) 87–93
13. Golder, S.A., Donath, J.: Social roles in electronic communities. Internet Research **5** (2004) 19–22
14. Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M.: Finding social roles in wikipedia. In: Proc. of ACM iConference. (2011) 122–129
15. Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., et al.: Knowledge discovery and data mining: Towards a unifying framework. In: Proc. of ACM Conference on Knowledge Discovery and DAta Mining. Volume 96. (1996) 82–88
16. Simmel, G., Wolff, K.H.: The Sociology of Georg Simmel. Macmillan Publishing Co. (1950)
17. Faust, K.: Very local structure in social networks. Sociological Methodology **37** (2007) 209–256
18. Davis, J.: Sociometric Triads as Multivariate Systems. Journal of Mathematical Sociology **5** (1977) 41–59
19. Baum, A., Shapiro, A., Murray, D., Wideman, M.V.: Interpersonal Mediation of Perceived Crowding and Control in Residential Dyads and Triads. Journal of Applied Social Phychology **9** (1979) 491–504
20. Burt, R.S.: Detecting role equivalence. Social Networks **12** (1990) 83–97
21. Holland, P., Leinhardt, S.: An Omnibus Test for Social Structure Using Triads. Sociological Methods and Research **7** (1978) 227–256
22. Doran, D., Alhazmi, H., Gokhale, S.: Triads, Transitivity, and Social Effects in User Interactions on Facebook. In: Proc. of IEEE Intl. Conference on Computational Aspects of Social Networks. (2013) 68–73
23. Viswanath, B., Mislove, A., Cha, M., Gummadi, K.: On the Evolution of User Interaction in Facebook. In: Proc. of 2nd ACM Workshop on Online Social Networks. (2009)
24. Maniu, S., Abdessalem, T., Cautis, B.: Casting a web of trust over wikipedia: an interaction-based approach. In: Proceedings of the 20th international conference companion on World wide web, ACM (2011) 87–88
25. Clauset, A., Shalizi, C.R., Newman, M.: Power-Law Distributions in Empirical Data. Technical report, arXiv:0706.1062v2 [physics.data-an] (2009)
26. Lipsky, L.: Queueing Theory: A Linear Algebraic Approach. 2nd edn. Springer-Verlag (2009)
27. Moody, J.: Matrix methods for calculating the triad census. Social Networks (1998) 291–299
28. Batagelj, V., Mrvar, A.: A subquadratic triad census algorithm for large sparse networks with maximum degree. Social Networks **23** (2001) 237–243
29. Ahmed, N.K., Neville, J., Kompella, R.: Network sampling: from static to streaming graphs. arXiv preprint arXiv:1211.3412 (2012)
30. Leskovec, J., Faloutsos, C.: Sampling from Large Graphs. In: Proc. of ACM Conference on Knowledge Discovery and Data Mining. (2006)
31. He, X., Zha, H., Ding, C.H., Simon, H.D.: Web document clustering using hyperlink structures. Computational Statistics & Data Analysis **41** (2002) 19–45

32. Gasch, A.P., Eisen, M.B.: Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. Genome Biol **3** (2002)
33. Jackson, J.E.: A User's Guide to Principal Components. John Wiley & Sons (2004)
34. Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining. Addison-Wesley (2006)
35. Medelyan, O., Witten, I.H., Milne, D.: Topic indexing with wikipedia. In: Proceedings of the AAAI WikiAI workshop. (2008) 19–24
36. Newman, M.: Networks: an introduction. Oxford University Press (2010)