

Persistent Roles in Online Social Networks

Matt Revelle (✉), Carlotta Domeniconi, Aditya Johri

George Mason University,
Fairfax VA 22030, USA

{reveille, carlotta}@cs.gmu.edu, ajohri3@gmu.edu

Abstract. Users in online social networks often have very different structural positions which may be attributed to a latent factor: roles. In this paper, we analyze dynamic networks from two datasets (Facebook and Scratch) to find roles which define users’ structural positions. Each dynamic network is partitioned into snapshots and we independently find roles for each network snapshot. We present our role discovery methodology and investigate how roles differ between snapshots and datasets. Six persistent roles are found and we investigate user role membership, transitions between roles, and interaction preferences.

1 Introduction

Online networks rely extensively on user contributions and participation for their vibrancy. This requires that users perform certain activities and take on specific roles within the network. In this paper we take a distinct approach to identify latent role behaviors which persist over time by examining interaction patterns and structural positions of users. Our approach provides a novel way of understanding latent mechanisms that underlie the structure and processes of dynamic networks.

Role discovery has been applied to many networks [22] and incorporated into static network models [27]. Despite the prevalence of role discovery methods and applications, no experiments have been presented that show the existence of persistent roles derived directly from data. While network-specific roles are useful for many purposes, identifying a set of roles which commonly occur in online social networks enables new methods for comparative analysis which emphasize relationships between roles.

In this paper, we present a methodology for discovering and tracing persistent roles over time. We discover roles for 26 network snapshots of online social networks from two datasets (Facebook and Scratch). These roles are found to persist both within and between the network snapshots from both datasets. We then conduct a summary analysis to demonstrate how roles may help interpret network structure by considering role membership, transitions between roles, and interaction preferences.

In our experiments, we discover six roles from the networks and show these roles are both distinct from one another and occur in every network from both datasets. These roles are: *popular*, *friendly*, *explorer*, *reciprocated*, *community*

member and *active-community member*. While the discovered roles are common to both datasets and persist over time, we find the relationships between roles may differ. These findings suggest common roles shared among social interaction networks are useful for modeling and comparing networks.

2 Related Work

2.1 Role Discovery

An overview of role discovery approaches is provided in [21] which discusses graph-based, feature-based, and hybrid definitions of roles and methods for their discovery from graph and node-attribute data. They show that feature-based roles are more flexible and capable of capturing more complex roles. A framework for feature-based role discovery is introduced and discusses classes of approaches for role feature construction and role assignment.

The use of non-negative matrix factorization (NMF) for discovering node roles was introduced in [9]. In that paper, the authors use a method [10] to generate features which aggregate various per-node structural attributes. This node-attribute matrix is then decomposed using NMF and the resulting basis vectors correspond to node roles in the network. Later work adds additional constraints to NMF which can be used to specify expectations of sparsity or diversity of the roles [7]. The work in our paper differs as we discover persistent roles across datasets and time using independent decompositions of network snapshots.

Other work [5, 30] uses role-labeled nodes to identify the roles of unlabeled nodes. However, the roles in their work are not defined in terms of structural positions in the network but rather functional occupations in an organization (e.g., roles held in technology companies: research & development, executives, and human resources). That is, the roles are defined in terms of domain knowledge and non-structural node features. The authors then introduce a classifier for these functional roles which incorporates information derived from the network structure.

Aside from identifying patterns of structural positions, roles have also been used in the context of information cascades to identify groups of nodes which have similar influence and blockage attributes [6].

A feature-based approach for automatic detection of user roles in online forums is presented in [4]. Their method uses principal component analysis (PCA) and agglomerative clustering of feature profile data to find roles; where each cluster corresponds to a role. Another feature-based approach using a mixture model of roles is presented in [27]. Nodes are first clustered using node features derived from the network structure and then a qualitative assignment of nodes to roles follows.

Role discovery has been used to assist in creating compact representations of networks. In [25] a method is introduced for generating multi-resolution maps of networks by constructing a hierarchy of roles defined with regular equivalence. The different levels of the hierarchy are used for different resolution maps.

2.2 Network Models with Roles

In [22] the role membership for a series of network snapshots are found and analyzed and the roles are used to construct a transition model of role memberships. Every node in every snapshot is represented as a mixed membership of roles. This mixed membership may change over time and a transition model captures the likelihood of transitioning between roles. Their method assumes roles are stationary and uses the same set of basis vectors (roles) for every snapshot rather than directly estimate roles from each snapshot. The authors suggest roles may generalize over time and across datasets, but do not provide support for this statement. To our knowledge, this paper is the first to present evidence of common, persistent roles derived directly from data.

Some models which incorporate roles do not distinguish between node features derived from network structure and those external to the graph. In [28], a probabilistic model which incorporates node features as dependent on latent factors (roles) is introduced. While these features could be derived from network structure as described in [21], the experiments performed in [28] only include external features such as document terms and voting counts. The network topology is ignored.

Communities provide extra structural information which can benefit role discovery. In both [23] and [8], communities are simultaneously detected with roles. Roles are used as latent factors of which node attributes are dependent.

Finally, [14] and [29] add roles to topic models where authors may take a role when generating a document and the topic of the document is dependent on the author’s role.

3 Discovering Persistent Roles

We aim to find roles which best characterize the nodes in a network. The network datasets we consider in this paper are dynamic networks which include timestamped, directed *interactions* between node pairs. Each interaction represents a single action such as one user messaging another. As our primary goal is to identify persistent roles over time, we will partition the dynamic network $\mathcal{D} = (\mathcal{N}, \mathcal{E})$ into snapshots, \mathcal{S}_t for each timestep t . The original edges \mathcal{E} are timestamped, directed interactions between node pairs and only edges occurring at timestep t , \mathcal{E}_t , are included in snapshot $\mathcal{S}_t = (\mathcal{N}_t, \mathcal{E}_t)$. The edges in \mathcal{E}_t are converted from individual interactions to directed, weighted edges, where the edge weight is the total number of directed interactions occurring between the nodes in \mathcal{S}_t . Nodes \mathcal{N} are derived from the edges \mathcal{E} and all nodes present at timestep t , \mathcal{N}_t , participate in at least one edge in \mathcal{E}_t .

3.1 Temporal Network Snapshots

The snapshots we construct are non-overlapping and each snapshot \mathcal{S}_t spans the same length of time, known as the observation window Ω . The structure of

network snapshots are defined by the activity which occurred within the observation window, thus there is no accumulation of inactive edges. The observation window Ω is calculated so that most time deltas δt_{ij} between interactions of any two nodes i and j are smaller than Ω . Specifically, we find the average time deltas $\langle \delta t_{ij} \rangle$ for each interacting node pair. The 90th percentile of all average time deltas is then used as Ω . We assume most connected pairs do not continually disconnect and reconnect and thus choosing an Ω which preserves most edges is appropriate. This methodology is described with more detail in [15, 19].

3.2 Role Feature Selection

Table 1: Node features

	Name	Description
1	<i>In-degree</i>	Count of incoming edges
2	<i>Out-degree</i>	Count of outgoing edges
3	<i>Weighted in-degree</i>	Count of incoming interactions
4	<i>Weighted out-degree</i>	Count of outgoing interactions
5	<i>Reciprocity</i>	Ratio of reciprocated edges over all outgoing edges
6	<i>New activity count</i>	Count of new outgoing edges
7	<i>Social strategy</i>	Ratio of new outgoing edges over all outgoing edges [15]
8	<i>Betweenness centrality</i>	Number of all shortest paths which pass through the node
9	<i>PageRank</i>	PageRank measure of centrality [17]
10	<i>Weighted PageRank</i>	Weighted variant of PageRank
11	<i>Transitivity</i>	Probability any two neighbor nodes are connected (local clustering coefficient) [26]
12	<i>Weighted transitivity</i>	Weighted variant of transitivity [2]

From the network snapshots we find D structural and behavioral features ($D = 12$ for our experiments) for all $n \in \mathcal{N}_t$ nodes and construct a matrix of node attributes $\mathbf{X}_t \in \mathbb{R}^{D \times N_t}$. The complete list of features used is shown in Table 1. Most of the features listed in Table 1 have common definitions, a few do not. The *new activity count* is computed for each node as the difference of the set of nodes reached from outgoing edges at the current snapshot \mathcal{S}_t and the set of nodes reached from outgoing edges at the previous snapshot \mathcal{S}_{t-1} . Similarly, *social strategy* is a ratio of the count of new outgoing edges (outgoing edges at snapshot \mathcal{S}_t) that did not exist at the previous snapshot over the total number of outgoing edges for the given node at snapshot \mathcal{S}_t , $\frac{\text{num. of new outgoing edges}}{\text{num. of all outgoing edges}}$. Users with a higher social strategy value tend to prefer making new connections (*social explorer*, or simply *explorer*) rather than preserve older connections (*social keeper*) [15].

These features were selected to enable the representation of the unique structural and behavioral patterns which may exist in online social networks which include individual, timestamped interactions. For example, while in-degree (count of incoming edges) captures popularity, the weighted in-degree (count of incom-

ing interactions, e.g., in Facebook, number of incoming wall comments) captures the overall level of incoming activity for the target node. Features such as *transitivity* encode information about a node’s neighborhood while *betweenness centrality* and *PageRank* capture global information about the node’s position in the network. The *reciprocity*, *new activity count*, and *social strategy* pertain to interaction behaviors.

3.3 Role Discovery and Membership

To find roles, a decomposition of a node-attribute matrix is performed and the resulting basis vectors are the discovered roles. We use non-negative matrix factorization (NMF) [13] for this task. The role vectors contain values corresponding to each feature which can be used to characterize the role — features with higher values are more characteristic of the role. For example, a role with a large in-degree might be labeled as popular.

NMF decomposes a matrix $\mathbf{X} \in \mathbb{R}^{D \times N}$ into a basis matrix $\mathbf{U} \in \mathbb{R}^{D \times L}$ and a coefficient matrix $\mathbf{V} \in \mathbb{R}^{L \times N}$, where L is the factorization rank of the decomposition $\mathbf{X} \approx \mathbf{UV}$. Each of the L columns of the basis matrix \mathbf{U} are the basis vectors or factors (roles) and the N columns of the coefficient matrix \mathbf{V} are the coefficient (weight) vectors which explain how each observation \mathbf{x}_i is represented as a mixture of roles.

NMF is independently run on the matrix of node attributes for each snapshot \mathbf{X}_t with the same parameters. We use the standard Euclidean update equation and Frobenius cost function. We use non-negative double singular value decomposition (NNDSVD) [3] to initialize NMF. This helps NMF converge faster and introduces a bias for sparse factors (roles). We do not expect roles will have non-zero values for all features as we assume roles are a parts-based representation [12] of node attributes. Each role is characterized by a subset of all available features.

3.4 Model Selection

A critical parameter of NMF is the factorization rank L . The common methods for selecting the rank value include: MDL [20], AIC [1], and error curves [16]. We initially tried to use MDL but found model size dominated the description length and resulted in the selection of low-performing models.

Recent existing work on role discovery with NMF [9, 22] used MDL and we attempted to use the same MDL function definition. Unfortunately, it appears the function does not appropriately balance between the model size and error for our datasets. We found that in all cases, the model with the lowest MDL had the smallest rank possible (for NMF with NNDSVD), $L = 2$.

We inspected the error curves, shown in Figure 1, and found that $L = 2$ results in a relatively large error. These curves were computed by calculating the root-mean-square error (RMSE) between the actual data \mathbf{X} and corresponding NMF approximation \mathbf{UV} . Instead of MDL, we elected to use the knee of the error curve to estimate the rank. As shown in Figure 1, networks across both

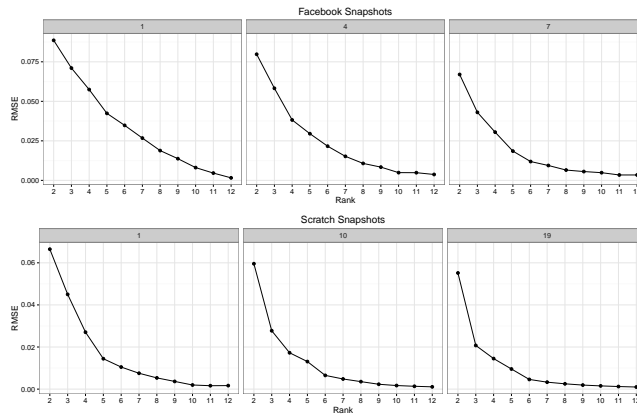


Fig. 1: Error curves for the first, mid, and final network snapshots in Facebook (top) and Scratch (bottom).

datasets had a similar error curve. Ranks $L = 5$ and $L = 6$ correspond to the knee point for most of the curves, and therefore are appropriate choices. Rank $L = 6$ is used for the factorization of all networks in our experiments.

3.5 Tracking Roles

Given T snapshots and node-attribute matrices for each snapshot \mathbf{X}_t , $t = 0 \dots T - 1$, NMF is used to perform the approximate decomposition $\mathbf{X}_t \approx \mathbf{U}_t \mathbf{V}_t$. Recall the basis matrix \mathbf{U}_t corresponds to role features and the coefficient matrix \mathbf{V}_t corresponds to role membership weights for each user. We hypothesize that roles may persist over time and need to verify whether the same roles do occur in consecutive basis matrices; i.e., do roles from \mathbf{U}_t appear in \mathbf{U}_{t+1} .

This role tracking is performed by measuring the similarity of every pair of role vectors between consecutive snapshots $\{\mathbf{u}_t^i \times \mathbf{u}_{t+1}^j \mid i, j \in 1 \dots L\}$. We use cosine similarity to evaluate the pairs and ensure that each role in snapshot t maps to only one role in snapshot $t + 1$ (the mapping is injective). We use a threshold value (0.75) to determine whether a pair matches. That is, if $\text{sim}(\mathbf{u}_t^i, \mathbf{u}_{t+1}^j) > 0.75$ then the pair of role vectors match. In practice, we find most matching pairs in our data have a cosine similarity greater than 0.9.

4 Data

We use two datasets of timestamped, directed interactions to construct dynamic networks and 26 network snapshots. The first dataset is a collection of Facebook wall posts [24] available from KONECT¹. In Facebook, users may post on each other’s wall and these posts are typically comments, photos, and web links. Each

¹ <http://konect.uni-koblenz.de/networks/facebook-wosn-wall>

of these posts is recorded as an interaction with a source user (the post author), a destination user (the owner of the wall), and a timestamp.

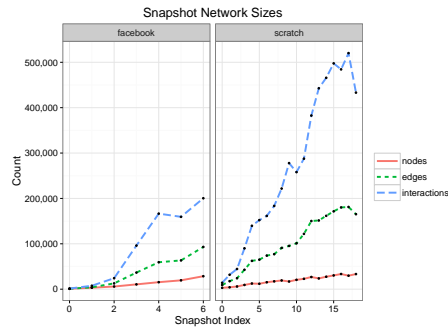


Fig. 2: Number of nodes, edges, and interactions over time in the Facebook and Scratch networks.

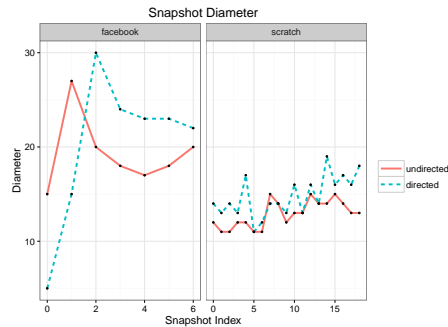


Fig. 3: Network diameter over time in the Facebook and Scratch networks.

The second dataset is a collection of Scratch project comments [18] extracted from a general Scratch dataset available from the MIT Media Lab website². Scratch is an online social network and web application for writing and sharing software programs. Programming education is the primary objective of Scratch and many users are children and young adults. Scratch users write and share projects; comments may be made on each other’s projects. Similar to Facebook walls, project comments in Scratch serve the purpose of public communication between users.

In both datasets, the interactions are used to construct a dynamic network and then network snapshots. The snapshots are constructed using the methodol-

² <https://llk.media.mit.edu/scratch-data>

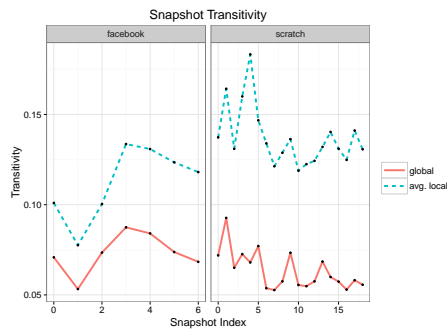


Fig. 4: Global and average local transitivity (clustering coefficient) over time in the Facebook and Scratch networks.

ogy discussed in Section 3.1. Figures 2, 3, and 4 show how the size and clustering of the snapshots from both datasets vary over time. Note that both the Facebook and Scratch interaction networks are growing over time.

A node-attribute matrix is created for each network snapshot using the features described in Section 3.2. Attributes are normalized by min-max normalization with all values belonging to the interval $[0, 1]$.

5 Results

We use the roles found by decomposing the per-snapshot, node-attribute matrix \mathbf{X}_t to answer our research questions. First we demonstrate that a common set of six persistent roles are found in the series of network snapshots from both datasets. While the feature proportions of the roles is similar across datasets and over snapshots, the magnitudes of the vectors change. Correspondingly, the magnitudes of the coefficient vectors (role membership weights) differ between snapshots.

We resolve this issue by averaging the basis vectors (roles) across all snapshots and then using non-negative least squares (NNLS) [11] to find the optimal coefficient matrix for the data, given the averaged basis matrix. This normalizes the role memberships between snapshots and these membership values are used in the rest of the analysis. Note that since the original basis vectors for all network snapshots had high cosine similarity, the averaged basis vectors also have a high cosine similarity with every original basis vector.

5.1 Persistent Roles

We use the methodology discussed in Section 3.3 to find roles in each network snapshot from both datasets. Then we follow the methodology described in Section 3.5 to determine whether the discovered roles occur in all snapshots from each dataset. We find six roles in both datasets which persist over time and

perform a pairwise comparison of the sets of roles from each dataset. There is a one-to-one correspondence (bijection) of the two sets of six roles, using the same cosine similarity test as was used for testing the persistence of roles across consecutive snapshots. That is, the same set of six roles persist over time in both datasets. We note that several roles are dominated by a single feature which is not shared with any other role, this suggests a parts-based factorization of node attributes.

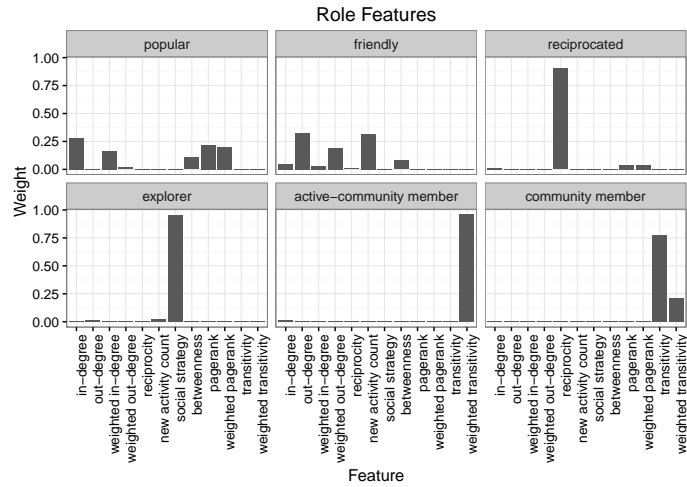


Fig. 5: Features for all roles, computed as average of role basis vectors from all network snapshots.

Figure 5 shows the discovered roles and their feature weights. The role names were selected according to the distinguishing features of the roles and we describe them here. The *popular* role is defined by the in-degree and centrality features while the *friendly* role has larger proportions in out-degree, weighted out-degree, and the number of new outgoing edges.

The *reciprocated* role is dominated by the reciprocity feature and captures the proportion of a node’s outgoing edges which are reciprocated by the receiver node. A node with perfect reciprocity would have a high membership weight in this role. The *explorer* role is dominated by the social strategy feature which indicates whether a node prefers to interact with new nodes rather than maintain existing relationships. We have observed that many nodes start as explorers when they first join the network.

The final two roles, *active-community member* and *community member*, capture the clustering of nodes. Active-community member is dominated by weighted transitivity which is similar to standard transitivity (local clustering coefficient) but accounts for the strength of the edge when calculating the coefficient. As we defined edge weight as the number of directed interactions between

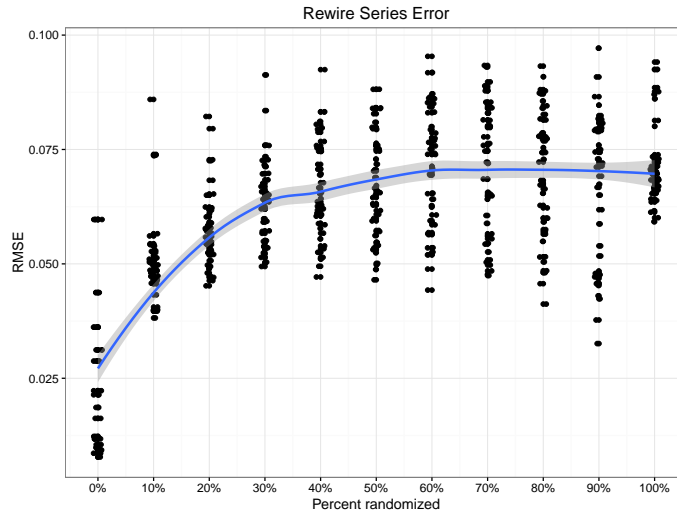


Fig. 6: Errors plotted for 50 series of randomly rewired networks.

a pair of source and destination nodes, a node with a high weighted transitivity coefficient is involved in an active community. In contrast, a node with a high unweighted transitivity coefficient simply participates in a densely-connected community and we cannot say anything about the activity of the community without further information.

5.2 Evidence of Role Dependence on Network Structure

We conduct an experiment with synthetic data to demonstrate the discovered roles capture patterns particular to the datasets. Fifty series of rewired networks were generated from networks in the original datasets. For each series, one of the snapshot networks was randomly selected. An increasing percentage of interactions in the network were removed and replaced with the same number of random interactions. Non-negative least squares (NNLS) is used to find the optimal role memberships (coefficient matrix) for each of the rewired networks.

The root-mean-square error between the actual data and the optimal approximation is calculated and Figure 6 shows the error increases as more interactions are randomly rewired. Thus our analysis supports the fact that the discovered roles reflect an intrinsic property of both social interaction networks, and not an artifact of the methodology used.

5.3 Role Membership

Using the persistent roles, we compare their distributions of role membership weights and check for correlations between roles. The role membership correlations (Spearman's coefficients) were calculated for every snapshot network,

however due to space constraints only the results for the final snapshot from Scratch is shown in Figure 7.

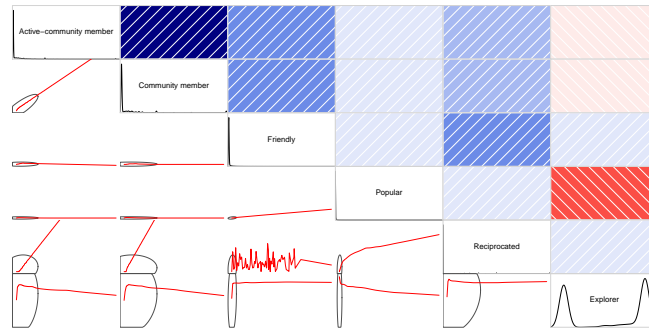


Fig. 7: Role correlations for the final snapshot from the Scratch dataset. The upper panels are colored to correspond to positive (blue) and negative (red) correlation. Darker shaded panels indicate larger correlation. The diagonal panels show the distribution of role membership weights. The lower panels show a confidence ellipse and smoothed line of the correlation.

The role membership correlations tend to be similar between all network snapshots in each dataset with one notable exception. Several correlations in early Facebook snapshots (popular and friendly, community member and friendly) shifted from having a negative correlation to a positive correlation. This change in Facebook may be due to the growth and sudden increase of activity after the first few snapshots.

5.4 Role Transitions

Nodes may be members of multiple roles and their role memberships may change over time. We visualize these transitions in Figure 9 for both the Facebook and Scratch datasets by identifying the top-5% nodes of each role for each network snapshot and draw a line between the roles of subsequent snapshots if nodes transition from one role to the other between those two snapshots. We select the nodes with the highest role membership weights as we expect them to be exemplary representatives of the roles. The height of the bars corresponds to the number of nodes with the role. A line is drawn between two roles if at least 10 users transitioned between the roles. The transition lines are sized according to the logarithm of the number of transitioning users. Since a user may share multiple roles, some transition lines merge and show users with multiple roles in common transitioning to a role in the next timestep. Figure 8 helps explain how to interpret the transition lines.



Fig. 8: A transition line from the *red* role to the *blue* role (left). A combined transition line from the *red* and *green* roles to the *blue* role (right). A combined line corresponds to transitioning users who belong in the top-5% of multiple roles in a single timestep.

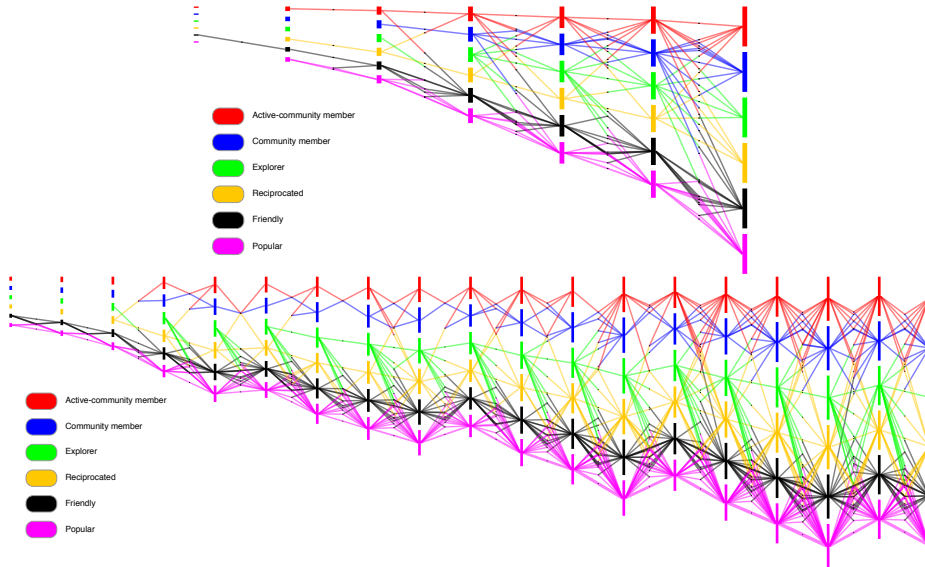


Fig. 9: The role transitions for the top-5% users in each role over all snapshots for Facebook (top) and Scratch (bottom).

As shown in [22], role membership of nodes may change over time and understanding these transitions allows us to construct predictive models. In this work, since a set of common roles has been identified, we can also perform comparative analysis of role transitions between the two datasets.

In both datasets, we see there are many transitions between *popular* and *friendly* roles as well as both community member roles. This is unsurprising as membership correlation is high for both pairs of roles. Further we note that neither *popular* nor *friendly* nodes ever transition to the *explorer* role. In contrast, users do transition from *explorer* to *popular* and *friendly*. This suggests that the most-popular users are less inclined to form new connections at the same rate as the top-5% *explorer* users.

There are also differences in the role transitions between the two datasets. In Facebook, we observe some community member nodes transition to the *explorer* role but this does not occur in Scratch. We hypothesize this may be attributed to the different uses of the social networks. While Facebook is a general social

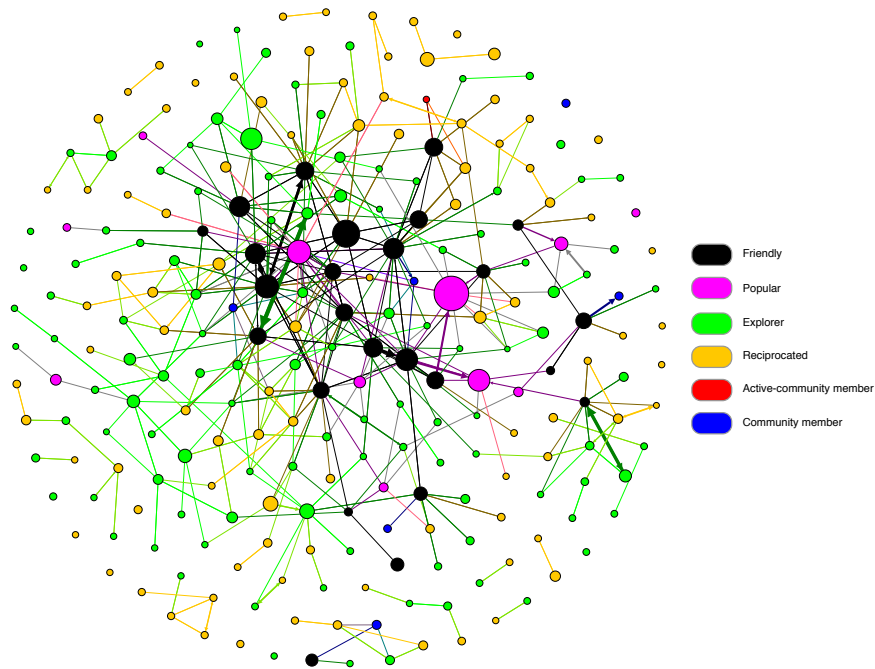


Fig. 10: A subgraph from a Facebook snapshot network. Nodes are colored by their primary role and sized according to their in-degree. Edges are sized according to the number of interactions they represent.

network, Scratch is used for teaching programming by schools and it is common for students in those classes to primarily only interact with other classmates.

5.5 Role Affinity

In this section we determine whether the persistent roles affect user link preferences. As the networks used in this study are directed, we consider both how roles impact the selection of nodes to interact with (outgoing) as well as how roles affect the attractiveness of some nodes (incoming). All nodes are assigned their primary role (the role with highest membership weight) for the role affinity analysis.

In Figure 10, we have colored nodes according to role and highlighted a subgraph for demonstration purposes. A standard force-directed layout algorithm was used to position the nodes. Note that while nodes with a higher in-degree tend to be either popular (magenta) or friendly (black), the friendly nodes have more outgoing interactions (larger outgoing edges). While friendly and popular roles reside in the core of the subgraph, explorer (green) and reciprocated (yellow) nodes appear on the periphery.

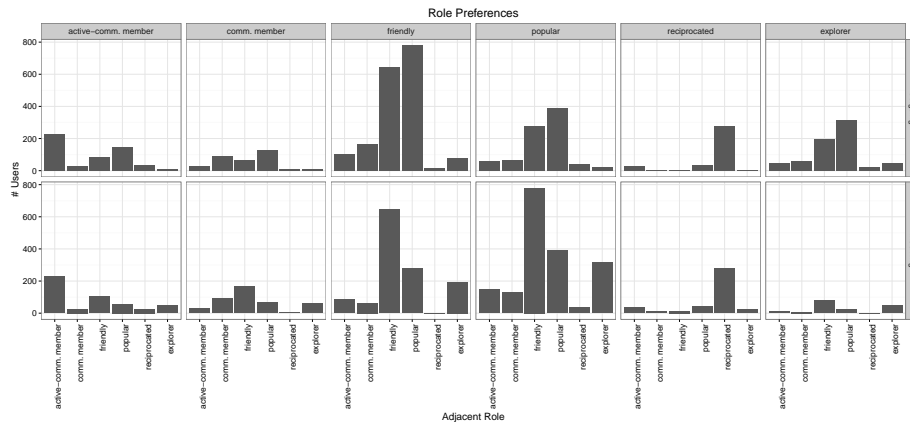


Fig. 11: The number of users with a primary role linked to/from other user roles. The column labels refer to the source node roles (for outgoing edges) and destination node roles (for incoming edges). The roles on the x-axis refer to the adjacent nodes.

We augment the network visualization with Figure 11 to present the exact counts of edges between roles. We note the lack of incoming edges to explorer nodes; evidence of this is also visible in the network of Figure 10.

6 Conclusion

User roles have become a critical component for improving our understanding of user interactions in online social networks. Persistent roles, shared between multiple datasets, enable a new comparative analysis method based on relationships between roles.

In this paper, we present a methodology for identifying persistent roles across time and datasets. Using this methodology, we find the same six user roles which capture distinct structural positions in 26 network snapshots from two online social networks. To our knowledge, this paper is the first to present evidence of persistent roles independently derived from multiple datasets. Beyond the discovery of persistent roles, we provide an analysis of the roles and show there are differences in role membership and interaction across the snapshots.

The findings presented in this paper will be leveraged in our future work to develop probabilistic models for the prediction of role membership and node attributes. We will also investigate the composition and evolution of communities viewed as interactions of roles.

Acknowledgement We appreciate the Lifelong Kindergarten group at MIT for publicly sharing the Scratch datasets. This work is partly based upon research supported by U.S. National Science Foundation (NSF) Awards DUE-1444277

and EEC-1408674. Any opinions, recommendations, findings, or conclusions expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

References

1. H. Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pages 199–213. Springer, 1998.
2. A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–3752, 2004.
3. C. Boutsidis and E. Gallopoulos. Svd based initialization: A head start for non-negative matrix factorization. *Pattern Recognition*, 41(4):1350–1362, 2008.
4. J. Chan, C. Hayes, and E. M. Daly. Decomposing discussion forums and boards using user roles. In *Proceedings of the 4th International Conference on Web and Social Media*, volume 10, pages 215–218, 2010.
5. Y. Cheng, A. Agrawal, A. Choudhary, H. Liu, and T. Zhang. Social role identification via dual uncertainty minimization regularization. In *Data Mining (ICDM), 2014 IEEE International Conference on*, pages 767–772. IEEE, 2014.
6. S. Choobdar, P. Ribeiro, S. Parthasarathy, and F. Silva. Dynamic inference of social roles in information cascades. *Data Mining and Knowledge Discovery*, pages 1–26, 2014.
7. S. Gilpin, T. Eliassi-Rad, and I. Davidson. Guided learning for role discovery (glrd): framework, algorithms, and applications. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 113–121. ACM, 2013.
8. Y. Han and J. Tang. Probabilistic community and role model for social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 407–416. ACM, 2015.
9. K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, and L. Li. Rolx: structural role extraction & mining in large graphs. In *Proceedings of the eighteenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1231–1239. ACM, 2012.
10. K. Henderson, B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, and C. Faloutsos. It’s who you know: graph mining using recursive structural features. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 663–671. ACM, 2011.
11. C. L. Lawson and R. J. Hanson. *Solving least squares problems*, volume 161. SIAM, 1974.
12. D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
13. D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.
14. A. McCallum, X. Wang, and A. Corrada-Emmanuel. Topic and role discovery in social networks with experiments on enron and academic email. *Journal of Artificial Intelligence Research*, pages 249–272, 2007.
15. G. Miritello, R. Lara, M. Cebrian, and E. Moro. Limited communication capacity unveils strategies for human interaction. *Scientific reports*, 3, 2013.

16. A. B. Owen and P. O. Perry. Bi-cross-validation of the svd and the nonnegative matrix factorization. *The annals of applied statistics*, pages 564–594, 2009.
17. L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.
18. M. Resnick, J. Maloney, A. Monroy-Hernández, N. Rusk, E. Eastmond, K. Brennan, A. Millner, E. Rosenbaum, J. Silver, B. Silverman, et al. Scratch: Programming for all. *Communications of the ACM*, 52(11):60–67, 2009.
19. M. Reville, C. Domeniconi, and A. Johri. Evidence of temporal artifacts in social networks. In *Proceedings of the 6th International Workshop on Mining Ubiquitous and Social Environments (MUSE)*, pages 35–42. 2015.
20. J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.
21. R. Rossi, N. K. Ahmed, et al. Role discovery in networks. *Knowledge and Data Engineering, IEEE Transactions on*, 27(4):1112–1131, 2015.
22. R. A. Rossi, B. Gallagher, J. Neville, and K. Henderson. Modeling dynamic behavior in large evolving graphs. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 667–676. ACM, 2013.
23. Y. Ruan and S. Parthasarathy. Simultaneous detection of communities and roles from large networks. In *Proceedings of the second edition of the ACM conference on Online social networks*, pages 203–214. ACM, 2014.
24. B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the evolution of user interaction in facebook. In *Proceedings of the 2nd ACM workshop on Online social networks*, pages 37–42. ACM, 2009.
25. J. Wang, K. C.-C. Chang, and H. Sundaram. Network cartography: Seeing the forest and the trees. *arXiv preprint arXiv:1512.06021*, 2015.
26. D. J. Watts and S. H. Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.
27. A. J. White, J. Chan, C. Hayes, and B. Murphy. Mixed membership models for exploring user roles in online fora. In *Proceedings of the 6th International Conference on Web and Social Media*, 2012.
28. R. Yu, X. He, and Y. Liu. Glad: group anomaly detection in social media analysis. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 372–381. ACM, 2014.
29. W. X. Zhao, J. Wang, Y. He, J.-Y. Nie, J.-R. Wen, and X. Li. Incorporating social role theory into topic models for social media content analysis. *Knowledge and Data Engineering, IEEE Transactions on*, 27(4):1032–1044, 2015.
30. Y. Zhao, G. Wang, P. S. Yu, S. Liu, and S. Zhang. Inferring social roles and statuses in social networks. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 695–703. ACM, 2013.