

Mining Twitter for an Explanatory Model of Social Influence

Jan Hauffa, Benjamin Koster, Florian Hartl, Valeria Köllhofer, and Georg Groh

Technische Universität München, Department of Informatics, Boltzmannstr. 3, 85748 Garching, Germany, {hauffa,koster,hartlf,koellhof,grohg}@in.tum.de

Abstract. The large-scale availability of online communication data offers an opportunity to learn about social influence on the individual level. Starting from an abstract cognitive definition, we iteratively build a predictive model of social influence upon the principle of locality of influence, which implies the decomposition of observed behavior into resistance to influence, and influence received via direct and indirect exposure to others' behavior. After training the model on a 30,000 user dataset of the social network service Twitter, we find that direct exposure has much less explanatory value than expected, and sources of influence exhibit strong temporal variation. We identify two modes of communication on Twitter, differing in the manifestation of influence.

1 Introduction

Interpersonal social influence has long been a subject of research in the social sciences. A generally accepted definition is “change in an individual’s thoughts, feelings, attitudes, or behaviors that results from interaction” [10], but the nature of the process, by which an individual receives influence, remains under active research and debate. With the rise of online social network services (SNS), social interaction has become observable outside of constrained experimental settings and accessible to large scale data mining. Longitudinal interaction data makes changes in behavior visible, enabling inference about changes in people’s attitude and reasoning about the process that drives these changes. By analyzing communication data in large volume, we attempt to identify fundamental characteristics of social influence.

1.1 Influence in Social Networks

In a simple model of human cognition, the behavior of an individual is determined by an internal state, which is constantly updated by perception of the environment. Change of behavior in reaction to events in the environment is the most general form of influence. The internal state is not observable, but observing both the environment and the behavior of an individual enables inductive reasoning about their relationship, and by extension about the underlying cognitive processes. Inferences can be tested by applying them to the prediction

of future behavior. Social influence can be defined as the subset of updates to the internal state caused by interpersonal interaction, and its effect on future interactions.

From an outside perspective, the effects of social interaction and general perception cannot be separated, so any amount of data that can be gathered in a practical experiment will be insufficient for reasoning within this model. To make inference tractable, we introduce an assumption called *locality of influence*: The influence of behavior perceived in social context a on behavior produced in context b is proportional to the similarity of a and b . Local influences may override external influence, but the resulting change in behavior may also be limited to a particular social context.

Related concepts can be found in the literature: Latané’s [5] dynamic theory of social impact asserts that “[...] influence is directly proportional to the immediacy of the source of influence.” Immediacy is defined as a combination of variables, including “richness of the communication channels” and geospatial distance. Myers et al. [8] provide empirical support by attributing only 29% of information in a complete record of Twitter activity over one month to “external events and factors outside the network”. The role of local graph structure for information diffusion in social networks is discussed e.g. by Zhang et al. [15].

1.2 Related Work

The main difference between our work and other studies of social influence [12] is our goal of learning about the influence process. Instead of inferring an influence network from observed interactions, our model yields a network-wide rule for generating individual influence networks for each user, comparable to egocentric diffusion networks [15].

2 Data Acquisition

Characterizing the social influence process requires a large corpus of observed social interaction that is not restricted to a particular social group or subject matter. We build such a corpus by crawling Twitter, an online service focused on the exchange of short text messages (“tweets”) up to 140 characters in length, which are public by default. The only method of interaction is posting a tweet, and the only relation over the set of users is “ a follows b ”, whereby a subscribes to tweets sent by b . Following is asymmetric, and does not require confirmation by the *followee*. Each user has a personal news feed that chronologically aggregates the tweets sent by followees.

2.1 Crawling Twitter

The follower network was crawled using non-exhaustive breadth-first search (BFS), ignoring the direction of edges. Accounts younger than 10 days, with a degree greater than 25,000, or not posting in English were excluded, to avoid

spammers and mitigate the effect of “hubs”, e.g. celebrities, who connect otherwise distant parts of the network.

Crawling produced a longitudinal dataset of 358,342 users and their tweets, which was subsampled to 30,000 users by BFS traversal from the original starting point due to the computational complexity of subsequent processing. Table 1 compares the samples to the full Twitter follower graph of July 2009 [4]. The metrics confirm that BFS is biased towards high-degree nodes, but preserves the dissortative tendency of the graph, and improves data quality for our use case by yielding subgraphs that are more dense than the original graph by orders of magnitude.

Table 1. Metrics of the Twitter follower graph

$ V $	30,000	358,342	41,652,230
$ E $	3,825,022	151,463,754	1,468,365,182
avg. degree	255.086	845.441	70.506
degree SD	407.538	2083.733	2534.992
density	0.009	0.002	$> 10^{-5}$
clust. coeff.	0.105	0.080	0.001
assortativity	-0.154	-0.066	-0.076

2.2 Social Conventions of Twitter Communication

The originally intended use case for Twitter was posting brief “status updates”. When holding conversations over Twitter became more popular, the community reached consensus on social conventions, which were later adopted by Twitter and integrated into the UI:

@-mention Prefixing a user name with the ‘@’ sign anywhere in a tweet causes the specified user to be notified. Honeycutt and Herring [3] identify two main uses: Addressing a message to another user, and referencing a user in a message intended for a wider audience.

Reply Tweets starting with an @-mention are considered part of an ongoing conversation.

Retweet Reposting a received tweet under one’s own name extends its visibility. The usual way of attribution is prefixing the quoted tweet with “RT” or “via”, followed by @-mentioning the original author.

Among the 17 million tweets of the 30,000 user dataset, 46% are regular tweets, 36% contain at least one @-mention, and 18% are retweets. 77% of tweets containing @-mentions are explicit replies via the UI. 8% of replies are users replying to their own posts, presumably chaining related posts.

Addressivity is a property of communication in online social media. The sender of an *addressive* message explicitly designates one or more recipients, demonstrating awareness. *Non-addressive* messages are “broadcast” to an undisclosed group of people. For the purposes of this work, we treat regular tweets as non-addressive and replies as addressive, while tweets containing @-mentions are counted both as non-addressive and as addressed to each mentioned user. On average, 36% of a user’s tweets are addressive ($\sigma = 24\%$).

Given the conceptual differences between the two types of communication, it stands to reason that they are also different in terms of influence, so we analyze them separately. As retweeting has already been studied within the information diffusion framework, e.g. by Zhang et al. [15], we exclude retweets from the following experiments.

2.3 Data Sparsity

Certain characteristics of the dataset may cause a lack of data in an experimental setting. The first issue is the low information content of a single tweet, caused by the size limit of 140 characters, and the presence of elements with a primarily social function, e.g. @-mentions. The second issue is sparsity of the spatio-temporal distribution of tweets. When discretizing time into periods of equal length, and assigning non-addressive and addressive messages to the nodes and directed edges of the social network graph, respectively, not all of them will be *active*, i.e. have at least one associated tweet, in each period. For a period length of 14 days, on average 69.2% of nodes and only 0.9% of edges were active, while for a period length of 2 days, 48.5% of nodes and 0.2% of edges were active. The third issue is missing observations. On average, only 19% of a node’s first degree neighbors in the Twitter follower graph are present in the sample.

3 Data Representation via Topic Modeling

The most salient component of interaction on Twitter is unstructured text, so a suitable numeric representation has to be found. Given evidence that individual potential to exert influence depends on the topic of conversation [6], topic models appear to be an appropriate choice.

Latent Dirichlet Allocation (LDA) [11] represents each document in a collection as a probability distribution θ over T topics, which in turn are probability distributions ϕ over the set of unique words. The Author-Recipient-Topic model (ART) [7], designed for email messages, extends LDA by observed variables for the sender and one or more recipients. For each sender-recipient pair, it yields a *relationship-topic distribution* representing the messages sent along the corresponding social graph edge. ART assigns each word of a message to an individual recipient. For short messages like tweets, it is more fitting to assume that the message as a whole is addressed to all recipients. As a compromise, we choose a canonical sender-recipient pair for each tweet: The first @-mentioned user in an addressive tweet is the recipient, while the author of a non-addressive tweet

is both sender and recipient, yielding separate topic distributions for each mode of communication.

3.1 Parameter Estimation and Inference

The tweet text is subjected to domain specific tokenization and stop word removal. The number of topics T is arbitrarily set to 150; values of the other ART hyper-parameters are chosen according to best practices: β is set to 0.01 [11] to obtain a symmetric Dirichlet prior for ϕ , while α is determined in a data-driven way [13], allowing the prior of θ to be asymmetric. Exact estimation of the model parameters is intractable, so we approximate them via 2000 iterations of Gibbs sampling.

For predicting behavior and evaluating the prediction, it is necessary to subdivide the dataset along the time axis, and compute separate relationship-topic distributions for each period. To be comparable, these distributions need to refer to a single set of topics ϕ . After parameter estimation on the full dataset, relationship-topic distributions for arbitrary subsets of the original data can be computed by resampling, i.e. repeating the Gibbs sampling process with fixed ϕ , for which 200 iterations are sufficient.

After resampling, the sampler’s internal state can be used for fast approximation of aggregate relationship-topic distributions over groups of senders and recipients. The formula for estimation of θ [7] is adapted to sum over a set of senders S and recipients R , resulting in 1 for approximation of the aggregate distribution $\theta_{S,R}$, where $t = 1..T$ is the topic index, and $n_{i,j,t}$ the number of words in messages from i to j assigned to topic t .

$$\theta_{S,R,t} = \frac{\alpha_t + \sum_{i \in S} \sum_{j \in R} n_{i,j,t}}{\sum_{t'=1}^T (\alpha_{t'} + \sum_{i \in S} \sum_{j \in R} n_{i,j,t'})} \quad (1)$$

After fitting an ART model to Twitter data covering a certain time period, we partition that data into observation and evaluation periods of equal length, and separate addressive from non-addressive communication. For each of these four subsets, various relationship-topic distributions (θ^M in Table 2) are computed via resampling and aggregation.

4 The Social Content Influence Model

The Social Content Influence Model (SCIM) learns to express the content of future interactions in terms of observed past interactions. Its predictive accuracy serves as an indicator for the explanatory value of the learned parameters.

Ignoring all other cognitive or social processes, future behavior can be fully explained by the presence or absence of social influence, or equivalently as a combination of *inertia* and *exposure* to others’ behavior. If exposure is potential influence, then inertia is individual resistance to influence, a tendency not to

deviate from past behavior. Unobserved sources of influence exist outside of the studied social medium, but also within, due to sampling. Their effect on the observed network appears as indirect influence, i.e. correlated behavioral changes in non-incident nodes [2]. Analogously, we distinguish *direct* and *indirect exposure*. If person a interacts with b , the content of the interaction can be directly observed, but will also be partially reflected in the future interactions of b with others. Aggregating the behavior of a group smoothes over individual preferences, but preserves information about strong influence that equally affected every member. With the principle of locality, it follows that the aggregated behavior of people who are socially close to b reflects the behavior b is exposed to.

From the perspective of an individual node or node pair (*ego* and *alter*) connected by an edge, the social network can be viewed as a hierarchy of social circles of decreasing locality. To account for missing observations within the medium, we aggregate over a node’s *social neighborhood*. Among different definitions of neighborhood, we aim to identify those that capture indirect exposure equally well across the whole graph. Influence from outside the medium is approximated by the aggregate behavior of the whole network, which potentially reflects strong trends from other media. This tripartite view of the egocentric social network corresponds to the distinction between interpersonal, peer, and media influence in sociology [14].

4.1 Prediction

Given the observed topic distributions from two successive time periods, the prediction problem can be formulated as using information from the first period to make predictions $\hat{\theta}_i^{M,n,s}$ for each node i , or $\hat{\theta}_{i,j}^{M,a,s}$ for each edge from i to j , so that their Jensen-Shannon divergence (JSD) from the distributions $\theta_i^{M,n,s}, \theta_{i,j}^{M,a,s}$ (see Table 2) in the second period is minimal. The JSD belongs to the family of symmetrized Kullback-Leibler divergences, which are commonly used for comparing topic distributions [11]. When defining the prediction $\hat{\theta}$ as a finite mixture of observed topic distributions $\theta^k, k \in C$ 2, finding coefficients c that minimize the JSD is a convex optimization problem 3.

$$\hat{\theta}_{i,j} = \left(\sum_{k \in C \setminus d} c_k \theta^k \right) + c_d \theta^d \quad (2)$$

$$\operatorname{argmin}_{c, \theta^d} \sum_{i,j} D_{JS}(\hat{\theta}_{i,j}, \theta_{i,j}) + \lambda \cdot \sum_{i,j} \|\hat{\theta}_{i,j}\|_1 \quad (3)$$

subject to $0 \leq c_k, \theta_t^d \leq 1$ for $k \in C, t = 1..T$,

$$\sum_{k \in C} c_k = 1, \sum_{t=1}^T \theta_t^d = 1$$

The models for addressive and non-addressive communication differ only in the number of mixture components. Table 2 lists all 15 components, names the

subset of messages they are computed from, and defines the set of senders and recipients they are aggregated over, where applicable. Each component represents either inertia, indirect, or direct exposure at a particular level of locality (*scope*). The components at relationship scope only apply to addressive communication.

Table 2. Mixture components of the SCIM

definition	S	R	role	scope
$\theta_i^{M,n,s}$ non-addr. messages sent by i	$\{i\}$	V	inertia	personal
$\theta_i^{M,a,s}$ addr. messages sent by i	$\{i\}$	V	inertia	personal
$\theta_{i,j}^{M,a,s}$ addr. messages from i to j	$\{i\}$	$\{j\}$	inertia	relationship
$\theta_i^{N(i),a,s}$ addr. messages from i to neighbors			inertia	neighborhood
$\theta_i^{M,n,r}$ non-addr. messages received by i	$\{x \in V : i \text{ follows } x\}$	V	direct exposure	personal
$\theta_i^{M,a,r}$ addr. messages received by i	V	$\{i\}$	direct exposure	personal
$\theta_{j,i}^{M,a,s}$ addr. messages from j to i	$\{j\}$	$\{i\}$	direct exposure	relationship
$\theta_i^{N(i),a,r}$ addr. messages from neighbors to i			direct exposure	neighborhood
$\theta_j^{M,n,s}$ non-addr. messages sent by j	$\{j\}$	V	indirect exposure	relationship
$\theta_j^{M,a,s}$ addr. messages sent by j	$\{j\}$	V	indirect exposure	relationship
$\theta^{N(i),n}$ non-addr. messages sent by neighbors			indirect exposure	neighborhood
$\theta^{N(i),a}$ addr. messages sent by neighbors			indirect exposure	neighborhood
$\theta^{M,n}$ all non-addr. messages	V	V	indirect exposure	medium
$\theta^{M,a}$ all addr. messages	V	V	indirect exposure	medium
θ^d estimated from data			indirect exposure	medium

Computing a single set of scalar coefficients that minimizes the error sum implies the assumption that the influence process is dominated by global, instead of individual or topical characteristics. Component θ^d is estimated from the data, capturing all global effects of influence that are either not explicitly represented in the SCIM or not directly observable. It allows the model to attain a training error of 0 if the influence process does not have any individual characteristics. The ℓ_1 regularization promotes sparse predictions and thereby the sparsity of c and θ^d . Regularization factor λ is set to 0.001.

4.2 Construction of the Social Neighborhood

The social neighborhood $N(i)$ of node i is a node-weighted subgraph of the social network graph (V, E) , induced by an indicator function $I_i : V \rightarrow \{0, 1\}$ and a weight function $W_i : V \rightarrow \mathbb{R}^+$. The neighborhood mixture components $\theta^{N(i)}$ are weighted sums over particular relationship-topic distributions of the subgraph nodes: $\theta_{i,v}^{M,a,s}$ for $\theta_i^{N(i),a,s}$, $\theta_v^{M,n,s}$ for $\theta^{N(i),n}$, $\theta_{v,i}^{M,a,s}$ for $\theta_i^{N(i),a,r}$, and $\theta_v^{M,a,s}$ for $\theta^{N(i),a}$.

We consider seven indicator and 25 weight functions. One family of indicators defines the neighborhood of i as the set of all nodes with a maximum distance of either one or two from i , either in the follower graph or the graph induced by addressive communication. The second family finds dense subgraphs of the undirected graph of reciprocal following, either by randomly selecting a maximal

clique containing i , or applying the clique percolation method ($k = 5$) [9] or edge clustering [1], and taking the union of the communities i is member of.

A basic weight function assigns uniform weight to all neighborhood nodes j . More complex functions derive the weight from structural properties of the social network graph (both local, such as the in-degree of j , and global, e.g. PageRank), from community structure (e.g. the number of shared communities of i and j), or from the communication behavior of j (e.g. how often j is retweeted).

5 Experimental Evaluation

The *basic prediction experiment* is defined as follows: First, a candidate set of either edges or nodes is built, depending on the type of communication to be analyzed. Candidates have to be active in both the observation and the evaluation period. The set is split randomly into training and test set of equal size, then parameter estimation and evaluation are performed.

This basic experiment is repeated, testing all combinations of four experiment parameters: The *observation date* marks the end of the observation and the beginning of the evaluation period. Three equidistant dates within eight weeks were chosen, April 20, May 4, and May 18 2012, aiming to test the temporal stability of the model. The length of the observation and evaluation period (*time period length*) needs to match the speed of conversation flow. We test periods of 14, 5, and 2 days, falling back to an extended period of 14 days if there is no activity. The *relationship type* is only relevant for addressive communication. It controls whether or not a needs to follow b for the edge from a to b to be considered. The last parameter is the choice of *social neighborhood*.

The SCIM is compared to three baseline predictors to verify that it captures non-trivial information about the influence process. The first predictor draws randomly from a Dirichlet distribution $\text{Dir}(\alpha)$ with α taken from the ART. The second predictor outputs the mean of $\text{Dir}(\alpha)$, which is the relationship-topic distribution the ART would produce in the absence of data. The third predictor outputs the relationship-topic distribution of the observed behavior, effectively a model of influence fully driven by inertia.

The experiment results are filtered to improve interpretability. Two restricted variants of the SCIM are introduced specifically to assess the utility of the coefficients and the neighborhood definitions. In the first variant, coefficients are uniform ($c_{1..|C|} = 1/|C|, c_d = 0$), while in the second variant all neighborhoods are empty. Any neighborhood definition that does not outperform these variants or the baselines across all combinations of experiment parameters is discarded.

To determine the experiment parameters' effect on prediction accuracy, we propose an ANOVA design, where the choice of neighborhood is a repeated measurement (including the baseline predictors for reference), and the remaining parameters are between-subject factors. The candidate sets are constructed and assigned to the experiments accordingly. All pairs of neighborhood definitions are tested post-hoc for significant differences in mean prediction error with Tukey's HSD test. The results can be expressed as homogeneous subsets of neighborhoods

with equivalent performance. After ranking them by mean error, the mixture coefficients of the best-performing subset are analyzed via descriptive statistics.

5.1 Results

43.4% of experiments for non-addressive, and 91.5% for addressive communication are filtered out. ANOVA is performed with a per-group sample size of 238. For both types of communication, there are significant interaction effects ($\alpha = 0.01$) involving neighborhood definition, observation date and time period length. This indicates that the amount of indirect influence captured by some or all of the neighborhood definitions varies over time, possibly related to the temporally irregular activity of users (Section 2.3). An interaction between neighborhood and time period length indicates that subgraphs differ in speed of information flow.

For non-addressive communication, there is a significant effect of time period length, with longer time periods improving the accuracy, but this effect may already be fully explained by the higher-order interactions. There is no significant effect involving the relationship type, so the existence of a follower relationship does not appear to affect the perception of addressive messages. For both types of communication, the choice of neighborhood is significant. Tukey’s test yields a high number of overlapping homogeneous subsets, but isolated baseline predictors. The lack of clustering limits the explanatory value of the best subsets.

The subset for non-addressive communication contains neighborhoods built by three indicator functions: First, communities found by edge clustering are given uniform weight, which implies that follower communities reflect indirect influence to a degree that is difficult to improve by weighting. Second, followers with a path distance of up to two, weighted with the number of shared followees or communities, also hint at the importance of cohesive social groups. Third, followers of distance one are paired with weights based on similarity of users or their message content, promoting homogeneous neighborhoods.

The neighborhoods in the best subset for addressive communication are built by a single indicator function, followers with a distance of up to two. Weights are mostly similarity-based and include the number of shared followees and the similarities of both kinds of communication.

Figure 1 compares the mean prediction error of the best subset to the baseline predictors. The SCIM outperforms all baselines, with a 10% improvement over the best performing baseline for non-addressive, and 28% for addressive communication. The lower error of the Dirichlet mean baseline predictor in case of addressive communication reflects the spatio-temporal sparsity discussed in Section 2.3.

Figure 2 shows the mixture coefficients as leaves of a tree, with the parent nodes representing either role or scope as listed in Table 2. Line width is proportional to the coefficient mean across the best subset, while the color corresponds to the ratio of mean and standard deviation: The darker, the less affected is the coefficient by the experiment parameters. Both addressive and non-addressive communication are strongly driven by inertia, but the predictive value of direct

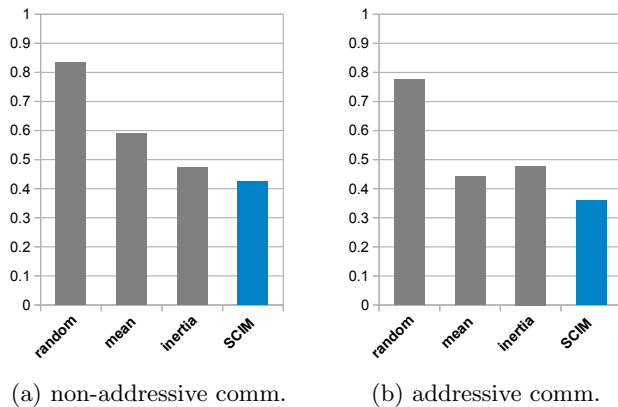


Fig. 1. Mean error of the SCIM and the baseline predictors

exposure is unexpectedly low, contradicting the principle of locality. The value of indirect exposure from the neighborhood is as expected, while the high value of the data-driven component θ^d suggests the existence of patterns of indirect influence not covered by the SCIM. Communication is mainly influenced by other communication of the same type. Components aggregating the relationship-topic distributions of a large number of users are generally of low predictive value.

6 Discussion

We report two main results: First, a novel point of view on the question whether Twitter is a social network, or a bipartite network of content producers and consumers [4]. A major difference to other social media is the high volume of non-addressive communication. Messaging behavior of individuals is highly variable, with the proportion of addressive communication having a one-SD range of 12% to 60%. The difference between the two modes of communication is visible in the influence process: Non-addressive communication is more resistant to influence, so the more stable communication behavior can be exploited by longer observation periods. Users are influenced in their non-addressive communication by their edge communities, while their addressive communication receives influence from a larger set of neighbors, weighted by similarity. In effect, the Twitter social network is a product of the follower network, which governs the flow of non-addressive communication, and the implicit network formed by addressive messaging.

Second, future behavior can be predicted to a certain extent from local sources of information, which the SCIM learns to exploit. However, our results do not fully confirm the decomposability of social influence into inertia, direct, and indirect exposure, which follows from the principle of locality. The low ex-

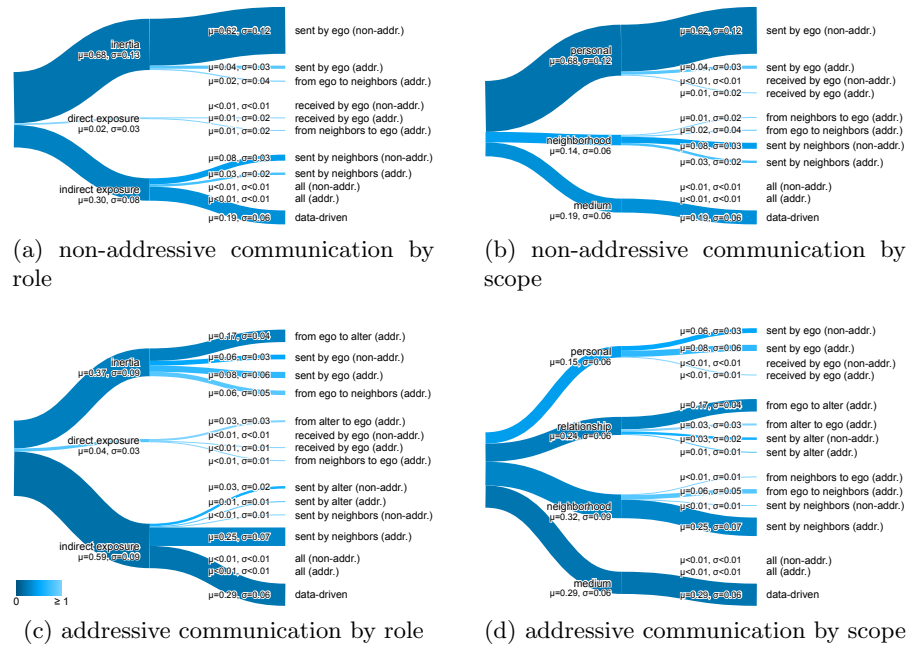


Fig. 2. Mixture coefficients of the SCIM experiments in the best subset

planatory value of direct exposure implies that locality is not sufficient on its own to explain why the SCIM is able to outperform the baselines: If interactions within and from outside the medium have similar potential for influence, observable interactions are responsible for just a fraction of the overall influence. Therefore it is important to exploit indirect influence, which allows information to cross the medium boundary. The best-performing neighborhood definitions favor nodes that are similar to the ego, and likely to be exposed to similar external influences.

Future work involves repeating the experiments on new datasets from different social media to test if our results apply to social interaction in general.

References

1. Ahn, Y., Bagrow, J., Lehmann, S.: Link communities reveal multiscale complexity in networks. *Nature* 466, 761–764 (2010)
2. Christakis, N., Fowler, J.: Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine* 32(4), 556–577 (2013)
3. Honeycutt, C., Herring, S.: Beyond microblogging: Conversation and collaboration via Twitter. In: *Proceedings of HICSS* (Jan 2009)
4. Kwak, H., Lee, C., Park, H., Moon, S.: What is Twitter, a social network or a news media? In: *Proceedings of WWW* (Apr 2010)

5. Latané, B.: Dynamic social impact: The creation of culture by communication. *Journal of Communication* 46(4), 13–25 (1996)
6. Liu, L., Tang, J., Han, J., Jiang, M., Yang, S.: Mining topic-level influence in heterogeneous networks. In: *Proceedings of CIKM* (Oct 2010)
7. McCallum, A., Wang, X., Corrada-Emmanuel, A.: Topic and role discovery in social networks with experiments on Enron and academic email. *Journal of Artificial Intelligence Research* 30, 249–272 (2007)
8. Myers, S., Zhu, C., Leskovec, J.: Information diffusion and external influence in networks. In: *Proceedings of SIGKDD* (Aug 2012)
9. Palla, G., Derényi, I., Farkas, I., Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435(7043), 814–818 (2005)
10. Rashotte, L.: Social influence. In: Ritzer, G. (ed.) *The Blackwell Encyclopedia of Sociology*, vol. 9, pp. 4426–4429. Blackwell (2007)
11. Steyvers, M., Griffiths, T.: Probabilistic topic models. In: Landauer, T., McNamara, D., Dennis, S., Kintsch, W. (eds.) *Handbook of Latent Semantic Analysis*, chap. 21. Lawrence Erlbaum (2007)
12. Sun, J., Tang, J.: A survey of models and algorithms for social influence analysis. In: *Social Network Data Analysis*, chap. 7. Springer (2011)
13. Wallach, H., Mimno, D., McCallum, A.: Rethinking LDA: Why priors matter. In: *Proceedings of NIPS* (Dec 2009)
14. Walther, J., Carr, C., Choi, S., DeAndrea, D., Kim, J., Tong, S., Van Der Heide, B.: Interaction of interpersonal, peer, and media influence sources online. In: Papacharissi, Z. (ed.) *A Networked Self*, chap. 1. Routledge (2010)
15. Zhang, J., Tang, J., Li, J., Liu, Y., Xing, C.: Who influenced you? Predicting retweet via social influence locality. *ACM Transactions on Knowledge Discovery from Data* 9(3), 25 (2014)