The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

COSINE: Community-Preserving Social Network Embedding from Information Diffusion Cascades

Yuan Zhang, Tianshu Lyu, Yan Zhang
Key Laboratory of Machine Perception (MOE)
Department of Machine Intelligence
Peking University, Beijing, China
{yuan.z, lyutianshu}@pku.edu.cn, zhy@cis.pku.edu.cn

Abstract

This paper studies the problem of social network embedding without relying on network structures that are usually not observed in many cases. We address that the information diffusion process across networks naturally reflects rich proximity relationships between users. Meanwhile, social networks contain multiple *communities* regularizing communication pathways for information propagation. Based on the above observations, we propose a probabilistic generative model, called COSINE, to learn community-preserving social network embeddings from the recurrent and time-stamped social contagion logs, namely *information diffusion cascades*. The learned embeddings therefore capture the *high-order user proximities* in social networks. Leveraging COSINE, we are able to discover underlying social communities and predict temporal dynamics of social contagion. Experimental results on both synthetic and real-world datasets show that our proposed model significantly outperforms the existing approaches.

Introduction

Network analysis has become an increasingly popular research area in the past decades with the emergence of online social media and social network sites. Network embedding is a fundamental problem in this area, aiming to learn a continuous representation of networks that can be used as input features for such downstream applications as visualization, node classification and link prediction (Tang et al. 2015).

Recent advances start from DeepWalk (Perozzi, Al-Rfou, and Skiena 2014) which uses random walks to generate node sequences as “sentences” and then exploits word2vec (Mikolov et al. 2013) developed in language modeling to obtain network embeddings. Tang et. al. (2015) argue that DeepWalk only captures the *second-order proximity* and further incorporate the *first-order proximity* into their proposed model LINE. Grover and Leskovec (2016) extend DeepWalk in another aspect, suggesting that a second-order random walk approach generates better learning samples. All of these approaches require prior knowledge of network structures. In many cases, however, we only observe when a node propagates information or becomes infected (e.g., retweeting and clicking like) but the structural connectivity is unknown (Gomez-Rodriguez, Balduzzi, and Schölkopf 2011). Besides, network structures, sometimes even when observed, only provide partial information about social dynamics. For instance, we cannot tell the interaction intensity between two users only from a who-follow-who network; a user might seldom contact some of her followees while interacting actively with others.

Bourigault et. al. (2014) first propose to learn social network embeddings from information diffusion cascades instead of network structures via a learning-to-rank approach. Later work (Kurashima et al. 2014; Bourigault, Lamprier, and Gallinari 2016) attempts to preserve temporal dynamics of information diffusion in social embeddings. Although they raise meaningful problems and provide inspiring preliminary solutions, most existing work fails to consider a very important notion in social networks, *communities*, meaning tightly knitted user clusters (Girvan and Newman 2002). The lack of consideration on communities may render it difficult to accurately deal with noisy social interaction logs involving highly volatile user behaviors (Hu et al. 2015).

More importantly, community structures in networks actually reflect the *high-order proximities*. Users in a same community often share similar opinions and behaviors due to fast social contagion within that community, and should therefore be considered to have close relationships.

In this paper, we propose a probabilistic generative model, COSINE (COmmunity-preserving SocIal Network Embeddings), to learn social network representations directly from more easily observed information diffusion cascades than network structures. As shown in Section 2, the nature of the diffusion process along with community structure regularization allows our approach to exploit not only the first-order and second-order proximities, but also, more generally, the high-order proximities between users. Leveraging COSINE, we are able to discover underlying social communities and predict temporal dynamics of social contagion.

Our approach jointly conducts metric learning (e.g., NetInf (Gomez Rodriguez, Leskovec, and Krause 2010), NetRate (Gomez-Rodriguez, Balduzzi, and Schölkopf 2011)) and multidimensional scaling (e.g., classic MDS (Cox and Cox 2001), LLE (Roweis and Saul 2000)). Zhou et. al. (2013) exploit l_1 and nuclear norm regularizations to avoid over-fitting issues when estimating metric matrices (i.e., social influence matrices) without prior information of network structures. In fact, we find that geometric constraints and clustering properties, associated with community structures, in the representation space can naturally provide sparsity and low-rank regularizations without additional complicated computations, respectively. The mutual reinforcement enables our model to outperform its pipelined variant carrying out these two procedures separately.

In summary, we make the following main contributions:

- We propose a novel generative model, called COSINE, for social network embedding from cascade data. To the best of our knowledge, COSINE is the first to exploit the information diffusion process regularized by community structures to preserve rich types of proximity relationships in social networks.
- We also design an efficient inference algorithm that guarantees linear scalability w.r.t. increasing input data.
- COSINE is evaluated with extensive experiments on both synthetic and real-world datasets. Results demonstrate its substantial performance improvement, robustness with limited training data, and application in understanding social contagion and interaction relationships in social networks.

Background: Information Diffusion Process

Our approach is built upon the continuous-time diffusion model for cascade data in social networks as in (Gomez-Rodriguez, Balduzzi, and Schölkopf 2011; Du et al. 2013). Compared with its discrete counterparts (Kempe, Kleinberg, and Tardos 2003) where information propagates iteratively in rounds, the continuous-time model is more appropriate in reality.

We assume that the diffusion process occurs over an unobserved underlying network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The process begins from a source user (node) u at time zero. The contagion is transmitted from the source user to its direct neighbors. Once infected, users continue to transmit the contagion to their respective neighbors. Each transmission through an edge, say (u, v), entails an independent random spreading time τ, we call in this paper the *local transmission time*, sampled from a distribution over time $f_{uv}(\tau)$. We assume a user cannot get infected twice, and thus the *global transmission time* t_{uv} for a user v to get infected is the earliest infection time over all possible paths from the source user u to the user v,

$$t_{uv} = \min_{p \in \mathcal{P}_{uv}} \sum_{(i,j) \in p} \tau_{ij}, \qquad (1)$$

where $\mathcal{P}_{uv}$ denotes the set of all paths from u to v.

Note that the diffusion process are not only determined by friendship-link distances (first-order proximity), but also relevant to the overall network structure due to group effects. In other words, even when user u and user v are not close neighbors, the information can still be quickly transmitted to v if they share similar neighbors (second-order proximity), or more generally, if they are in the same interaction-intensive *communities* (high-order proximities). For example, in Figure 1, the contagion originated from user 0 spreads to user 5 almost as fast as to its direct neighbor user 1 because they share common neighbors (i.e., user 3 and user 4); user 9 is expected to receive the contagion earlier than user 2 thanks to the densely connected community, although two-hops farther away from the source.

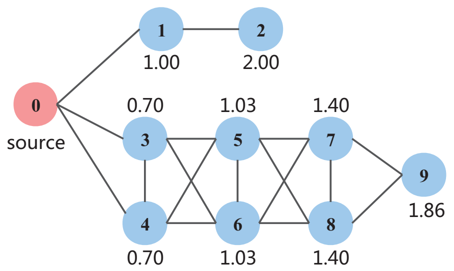


Figure 1: An illustrative example of information diffusion. User 0 is the source user. The numbers near other nodes indicate the expected global transmission time intervals where local transmission time intervals over all edges are sampled from the same distribution, $Exponential(1.0)$. We note that the information diffusion process regularized by community structures reflects rich proximity relationships between users.

This property naturally enables our diffusion-based approach to learn embeddings that capture the *high-order proximities* between users in social networks.

Proposed Model: COSINE

Model Framework

Figure 2 shows the probabilistic graphical representation of our model. The input is the diffusion cascades $\mathcal{C} = \bigcup_{u=1,\dots,N} \mathcal{C}_u$, where N is the number of users and $\mathcal{C}_u$ denotes the diffusion cascades originated from user u. Each cascade c includes a set of tuples, namely $\{(v, t_v^c)\}_{v \in \mathcal{U}_c}$, of infected users $\mathcal{U}_c$ and their infection timestamps (global transmission time) within an observation time window $(0, T_c]$. Here, we reset the "clock" to 0 at the start of each cascade and assume $T_c = T$ for each cascade c for simplicity.

Our goal is to learn a low-dimensional continuous representation $x_u \in \mathbb{R}^D$ for each user u, where D denotes the number of dimensions. Each dimension may act as a subset of social interaction channels. We use the squared Euclidean distances[1] (i.e., $d_{uv} = \|x_u - x_v\|^2$) in the low-dimensional space to represent mean values of the global transmission time distributions from which cascading timestamps are drawn. Thus, our approach can jointly model temporal dynamics of information diffusion and learn community-preserving embeddings "with mutual benefits".

Modeling Temporal Dynamics of Information Diffusion with Long-Tailed Distributions. Information diffusion

[1] We have tested several dissimilarity measures and this one offers a good compromise.

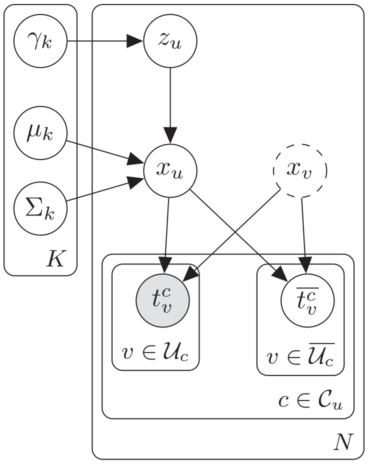


Figure 2: Probabilistic graphical representation of our model.

process reflects rich proximity relationships between users. However, data sparsity and volatile user behaviors make it difficult to model temporal dynamics of information propagation. Besides, we can only observe the "early stages" of the information diffusion processes within an observation window T. Normal distributions might be easily biased towards longer transmission time and sensitive to missing data (an "infected" user would be considered as "uninfected" in a cascade if her record were missing).

Therefore, we propose to exploit a long-tailed distribution, inverse-Gaussian distribution (a.k.a. Wald distribution) that is often used in survival data analysis, to model the sparse infection timestamps with large uncertainty. We refer the reader to Model Inference for more intuitions.

Specifically, supposing user u is the source in a cascade c, we let the squared Euclidean distance $d_{uv} = \|x_u - x_v\|^2$ in the representation space denote both the mean and variance parameters of the global transmission time distribution for user v, i.e.,

$$f(t_v^c) = \frac{d_{uv}}{\sqrt{2\pi}(t_v^c)^{3/2}} \exp\{-\frac{1}{2} \cdot \frac{(t_v^c - d_{uv})^2}{t_v^c}\}. \qquad (2)$$

Meanwhile, we regard timestamps $\{\overline{t_v^c}\}$ of those "survival" users who are not involved in a given cascade as missing data where the missing mechanism is $\mathbb{I}(\overline{t_v^c} > T)$ and use statistical methods to impute those missing timestamps. Compared with those leveraging alternative survival analysis methods (e.g., (Kurashima et al. 2014; Bourigault, Lamprier, and Gallinari 2016)), our approach is more efficient and scalable, especially when the lengths of cascades are relatively large.

Preserving Community Structure. There exist multiple dense subgraphs, communities, in social networks with close connection to information contagion. In our approach, we use the Gaussian mixture model (GMM) as priors on users' embeddings to preserve community structure in the given network. We assume that there are K communities, and that each user u belongs to exactly one community denoted by $z_u \in [1, ..., K]$ drawn from a multinomial distribution $\{\gamma_k\}$.

Table 1: Notations used in our paper.

Symbols	Descriptions
N, K, D	numbers of users, communities, dimensions
T	observation time window
$\mathcal{C}, \mathcal{C}_u$	sets of diffusion cascades and those originated from user u
$\mathcal{U}_c, \overline{\mathcal{U}_c}$	sets of infected and uninfected users in cascade c
z_u	latent community membership of user u
x_u	embedding vector of user u
d_{uv}	squared Euclidean distance between embedding vectors of user u and v, i.e., $\|x_u - x_v\|^2$
$t_u^c, \overline{t_u^c}, \widetilde{t_u^c}$	observed, unobserved, imputed timestamps of user u getting infected in cascade c
γ_k	prior probability / mixing weight of cluster k
r_{uk}	conditional probability user u in cluster k
μ_k, Σ_k	mean, variance matrix of cluster k
I_{max}	number of maximum iterations

Each community k is represented by a Gaussian component with mean μ_k and covariance matrix Σ_k that reflect the community's relative position and connectivity density, respectively.

Our model can be extended to capture overlapping communities with an unknown number of communities by changing GMM to hierarchical mixture models with Dirichlet priors (e.g., the one proposed in (Teh et al. 2006)) with little effort. Yet, this paper only focuses on the simple case for clarity.

We summary the notations used in this paper in Table 1.

Model Inference

Similar to (Yang, Tang, and Cohen 2016), we employ an alternating optimization approach for inference. First, we fix the embeddings (i.e., $\{x_u\}$) to update GMM parameters (i.e., $\{\gamma_k\}$, $\{\mu_k\}$, and $\{\Sigma_k\}$):

$$\gamma_k = \frac{N_k}{N}, \mu_k = \frac{\sum_u r_{uk} x_u}{N_k}, \qquad (3)$$

$$\Sigma_k = \frac{\sum_u r_{uk}(x_u - \mu_k)(x_u - \mu_k)^T}{N_k}, \qquad (4)$$

where $r_{uk} \triangleq P(z_u = k | x_u) = \frac{\gamma_k \mathcal{N}(x_u|\mu_k, \Sigma_k)}{\sum_{k'} \gamma_{k'} \mathcal{N}(x_i|\mu_{k'}, \Sigma_{k'})}$ and $N_k \triangleq \sum_u r_{uk}$. The community membership for each user u can be obtained by MAP estimation,

$$\hat{z}_u = arg \max_k r_{uk} = arg \max_k P(z_u = k | x_u). \qquad (5)$$

Then, we fix the GMM parameters and update the embeddings via a generalized EM method to fit the survival inverse-Gaussian timestamps in the observation window. In the E-step, we calculate the conditional expectations to impute "missing" timestamps $\{\overline{t_v^c}\}$ according to (Whitmore 1983),

$$\widetilde{t_u^c} = E[\overline{t_v^c} | \overline{t_v^c} > T] = d_{uv} \cdot \frac{H(d_{uv}^2/T, d_{uv})}{1 - H(T, d_{uv})}, \qquad (6)$$

where u is the source user of cascade c and H is the distribution function of the inverse-Gaussian distribution.

In the M-step, we first write down the expected complete data log-likelihood function to optimize,

$$\begin{aligned}\mathcal{L} = &\sum_{u=1}^{N}[\sum_{c\in\mathcal{C}_u}(\sum_{v\in\mathcal{U}_c}\log f(t_v^c) + \sum_{v\in\overline{\mathcal{U}_c}}\log f(\widetilde{t}_v^c)) \\ &+ \log\sum_{k=1}^{K}\gamma_{uk}\mathcal{N}(x_u;\mu_k,\Sigma_k)] \\ \geq &\sum_{u=1}^{N}[\sum_{c\in\mathcal{C}_u}(\sum_{v\in\mathcal{U}_c}\log f(t_v^c) + \sum_{v\in\overline{\mathcal{U}_c}}\log f(\widetilde{t}_v^c)) \\ &+ \sum_{k=1}^{K}\gamma_{uk}\log\mathcal{N}(x_u;\mu_k,\Sigma_k)] = \mathcal{L}'\end{aligned} \tag{7}$$

where the inequality is obtained thanks to log-concavity, i.e., $\log\sum_{k=1}^{K} r_{uk}\mathcal{N}(x_u;\mu_k,\Sigma_k) \geq \sum_{k=1}^{K} r_{uk}\log\mathcal{N}(x_u;\mu_k,\Sigma_k)$. We then employ gradient ascent to maximize the lower-bound $\mathcal{L}'$ of the log-likelihood by updating the embeddings $\{x_u\}$. The gradients are computed as,

$$\begin{aligned}\frac{\partial\mathcal{L}'}{\partial x_u} = &\sum_{c\in\mathcal{C}_u}\sum_{v\in\mathcal{U}_c}(\frac{d_{uv}-t_v^c}{t_v^c} - \frac{1}{d_{uv}})\cdot(x_v - x_u) \\ &+\sum_{v=1}^{N}\sum_{c\in\mathcal{C}_v|u\in\mathcal{U}_c}(\frac{d_{vu}-t_u^c}{t_v^c} - \frac{1}{d_{vu}})\cdot(x_u - x_v) \\ &+\sum_{c\in\mathcal{C}_u}\sum_{v\in\overline{\mathcal{U}_c}}(\frac{d_{uv}-\widetilde{t}_v^c}{\widetilde{t}_v^c} - \frac{1}{d_{uv}})\cdot(x_v - x_u) \\ &+\sum_{v=1}^{N}\sum_{c\in\mathcal{C}_v|u\in\overline{\mathcal{U}_c}}(\frac{d_{vu}-\widetilde{t}_u^c}{\widetilde{t}_u^c} - \frac{1}{d_{uv}})\cdot(x_u - x_v) \\ &+\sum_{k=1}^{K} r_{uk}\Sigma_k^{-1}(\mu_u - x_u),\end{aligned} \tag{8}$$

where $d_{uv} = d_{vu} \triangleq \|x_u - x_v\|^2$.

We briefly explain the intuition behind x_u update. It is influenced by both the difference between the distance and the expected global transmission time (the first four terms) and the consistency with the centroids of communities user u likely belongs to (the last term). The former is weighted by the inverse of transmission time. Thus, we emphasize smaller global transmission time intervals that would contain more information for preserving local network structures, and de-emphasize larger ones that are more noisy and unstable. We refer interested readers to (Luo et al. 2011) for a similar idea of network embedding. Besides, short distances are penalized by the term $-\frac{1}{d_{uv}}$ to avoid over-fitting.

Here, dealing with non-infected users is the most time-consuming part (Eq. (6) and the third to forth lines of Eq. (8)) due to sparsity of diffusion cascades. This computational burden can be reduced by the idea of negative sampling (Mikolov et al. 2013). In our case, we sample a non-infected user set $\overline{\mathcal{U}_c}$ twice[2] as large as that of the infected user set $\mathcal{U}_c$ for each cascade c.

We iteratively repeat the procedures above for a given number of iterations until convergence as shown in Algorithm 1.

Algorithm 1 Model Inference

```
Input: Diffusion cascades $\mathcal{C}$, number of users $N$, number
    of communities $K$, number of dimensions $D$, learning
    rate $\alpha$, maximum iterations $I_{max}$.
Output: Embeddings $\{x_u\}$, community memberships
    $\{\hat{z}_u\}$, GMM parameters $\{\gamma_k, \mu_k, \Sigma_k\}$
Randomly initialize $\{x_u\}$
for $i \leftarrow 1$ to $I_{max}$ do
    for $k \leftarrow 1$ to $K$ do
        Calculate $\gamma_k, \mu_k, \Sigma_k$ according to Eq. (3-4)
    end for
    for all $c \in \mathcal{C}$ do
        Sample non-infected users $\overline{\mathcal{U}_c}$
        for all $u \in \overline{\mathcal{U}_c}$ do
            Impute $\widetilde{t}_u^c$ according to Eq. (6)
        end for
    end for
    for $u \leftarrow 1$ to $N$ do
        Update $x_u$ by gradient ascent according to
        Eq. (8)
    end for
end for
for $u \leftarrow 1$ to $N$ do
    Calculate $\hat{z}_u$ according to Eq. (5)
end for
```

Time Complexity Analysis

In each iteration, we first update GMM parameters with complexity $O(KN)$. The E-step takes $O(\sum_c|\mathcal{U}_c|) = O(|\mathcal{C}|)$ to impute timestamps of sampled non-infected users (twice as many as the infected users in our case), while the following M-step takes $O(|\mathcal{C}|+KN)$ to update embeddings. Here, we use $|\mathcal{C}|$ to denote the size of input cascades, i.e., the number of all tuples contained in each cascade in $\mathcal{C}$. Overall, the complexity is $O(I_{max}(|\mathcal{C}| + KN))$, where I_{max} is the number of maximum iterations.

Therefore, when K and I_{max} are regarded as given constants, our inference algorithm scales linearly in terms of the size of the input data, i.e., the number of users and the size of diffusion cascades.

Experiments

In this section, we perform various experiments on both synthetic and real-world datasets to evaluate our proposed approach and conduct a case study of online social media sites crawled by Memetracker. We compare COSINE with the following state-of-art baseline methods:

[2] There is clearly a trade-off between accuracy and computational cost when determining the ratio of negative samples to positive samples. We here empirically choose "twice".

C-Rate, C-IC (Barbieri, Bonchi, and Manco 2013) leverage diffusion cascades to discover latent communities in a network-oblivious setting.

CDK (Bourigault et al. 2014) is a ranking-based network embedding method to represent users in such a latent vector space that information diffusion can be regarded as a heat diffusion process in that space.

NetRate (Gomez-Rodriguez, Balduzzi, and Schölkopf 2011) is a network inference method for estimating information transmission rates between users based on cascade data.

Pipeline is the pipelined variant of our model. We first estimate temporal infectivity between users, and then learn network embeddings similar to COSINE.

Synthetic Dataset

Data Generation. As in (Barbieri, Bonchi, and Manco 2013), we first generate four networks with known community structures by the widely used benchmark for community detection (Lancichinetti, Fortunato, and Radicchi 2008). The process of network generation is controlled by the following parameters: (i) number of nodes (1,500), (ii) average degree (10), (iii) maximum degree (200), and (iv) min/max community sizes (50/400). The four networks differ in the mixing parameters λ that control the fraction of edges of a node that go outside its community, ranging in [0.05, 0.1, 0.15, 0.2].

Then, we simulate information diffusion processes to generate synthetic diffusion cascades with an observation window $T = 4.0$. For each edge, the local transmission time intervals are sampled from a exponential distribution, while the transmission rate is sampled from a Gamma distribution with the shape parameter 2 and scale parameter 0.025.

Community Detection. Since the ground truth of communities are available in synthetic datasets, we directly compare COSINE's performance on community detection with other baseline methods using NMI (normalized mutual information) as the evaluation metric. We exploit GMM to obtain clusters as communities for CDK and Pipeline, and use the Metis package[3] (Karypis and Kumar 1998) to detect communities from the networks inferred by NetRate. Figure 3 shows that COSINE substantially outperform other baselines. Without the mutual regularization between social infectivity and community-preserving network embeddings, the pipelined approach exhibits worse performance than COSINE especially as the mixing parameter increases (i.e., community structures are less clear).

Figure 5 compares the visualization of different embedding approaches. We can see that COSINE not only exhibits better accuracy, but also guarantees more distinct boundaries between different communities.

Diffusion Prediction. We also evaluation our model in the task of diffusion prediction. We independently generate a set of test cascades and predict whether users are infected in each cascade in a given time window T' (here, we choose $T' = 1.5$). Because CDK only scores the likelihood of contagion by users' distance from the source user without a prediction threshold, we use the averaged AUC as the evaluation metric for a fair comparison. As shown in Table 2, COSINE demonstrates significant improvement than baseline methods. Community regularization enables COSINE and C-Rate to perform relatively better as the mixing parameter increases, while it is the opposite for other baselines that fail to consider community structures in networks due to over-fitting issues.

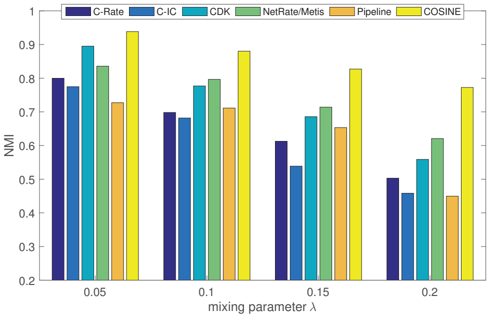


Figure 3: Community detection performance with different mixing parameters $\lambda = 0.05, 0.1, 0.15, 0.2$.

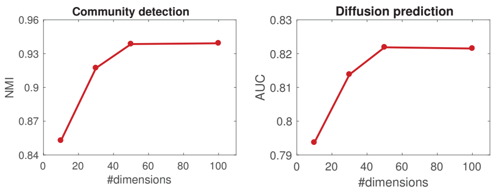


Figure 4: Sensitivity w.r.t. different numbers of dimensions.

Parameter Sensitivity. We also demonstrate the sensitivity w.r.t. the numbers of parameter dimensions D. As in Figure 4, larger dimensions brings better results until reaching saturation at around 50. Therefore, we set $D = 50$ in the other experiments when not otherwise specified.

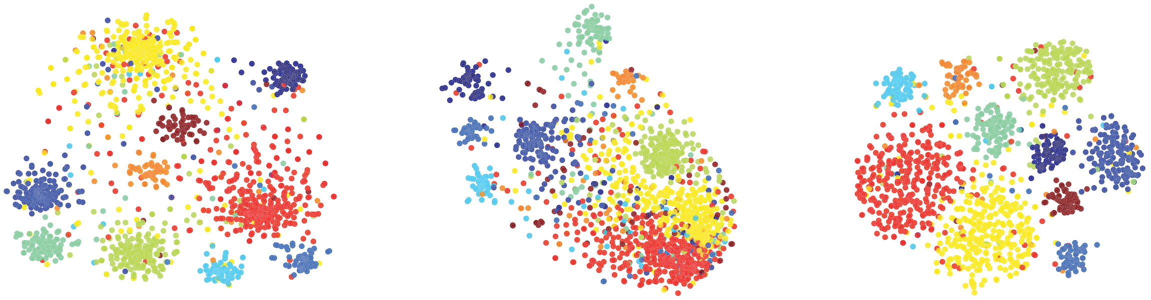
Real-world Dataset

We use the Memetracker dataset (Leskovec, Backstrom, and Kleinberg 2009) which tracks the diffusion of popular quotes and phrases, called "memes", across online media sites and blogs from August 2008 to April 2009. We extract the most active 1,000 media sites as "users" with 6,000 cascades that are split into halves for training and testing.

We investigate the predictive performance with different numbers of training cascades. As illustrated in Figure 6(a), COSINE demonstrates consistent improvement over other baseline methods especially when the number of training cascades is less than 2,000. This result suggests robustness of our model in the setting of sparse data as is often the case in social network analysis.

Case Study. Figure 7 visualizes communities of social media sites detected by COSINE with recognized community themes. Their relative positions and densities provide rich information on proximity relationships both at the indi-

[3] http://glaros.dtc.umn.edu/gkhome/metis/metis/overview



(a) CDK

(b) Pipeline

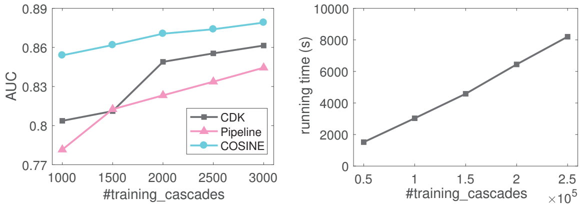
(c) COSINE

Figure 5: Visualization of the synthetic networks with mixing parameter $\lambda = 0.15$. Nodes are projected into the 2-D space using t-SNE (Maaten and Hinton, 2008) with the learned embeddings as input. Color of a node indicates the community membership.

Table 2: Diffusion prediction performance (AUC) on synthetic datasets.

λ	C-Rate	CDK	NetRate	Pipeline	COSINE
0.05	0.6883	0.8093	0.7877	0.7846	**0.8219***
0.10	0.6784	0.8132	0.7778	0.7882	**0.8347**†
0.15	0.7001	0.8114	0.7687	0.7717	**0.8425**‡
0.20	0.7025	0.7989	0.7560	0.7636	**0.8378**‡

Significantly outperform the best baseline method: $*(p < 0.05)$, $\dagger(p < 0.01)$, $\ddagger(p < 0.005)$



(a)

(b)

Figure 6: (a) Diffusion prediction performance on Memetracker with different numbers of training cascades. We omit inapplicable or uncompetitive (AUC < 0.8) baselines for better presentation. (b) Running time versus numbers of training cascades.

vidual level and the community level. For example, the communities of French sites are most "exclusive" with almost no interaction with other communities, while some individual political sites are embedded in the three communities of mainstream news in the middle or in the community of entertainment to its left, exhibiting different media styles.

We also list some typical domains and sample memes of four representative communities in Table 3. The community of Spanish sites contains top-level domains of multiple Spanish-speaking countries (e.g., Spain *.es*, Mexico *.mx* and Colombia *.co*) with less closely knitted representations in the 2-D space than its counterparts, the communities of French sites and German sites.

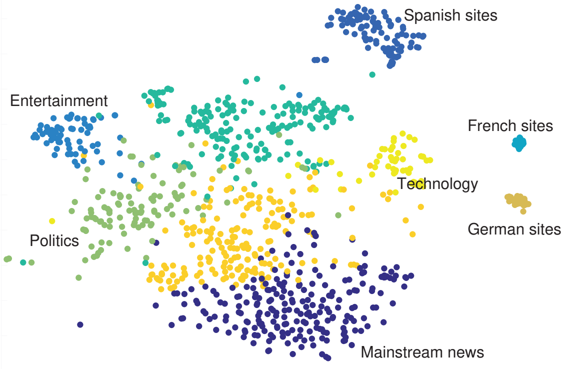


Figure 7: Visualization of Memetracker with COSINE.

Running Time. For the sake of comparison with all baseline methods, we only focus on relatively small datasets in experiments. Yet, our approach guarantees linear scalability with regard to increasing input data. Figure 6(b) plots the running time of COSINE versus the numbers of training cascades. The experiments are run on Intel(R) Xeon(R) E5-2620@2.00GHz with 64GB of RAM.

Related Work

Network Embedding. Network embedding recently attracts extensive research interests. In addition to such word2vec-based methods as DeepWalk (Perozzi, Al-Rfou, and Skiena 2014), LINE (Tang et al. 2015), GraRep (Cao, Lu, and Xu 2015) and node2vec (Grover and Leskovec 2016), Planetoid (Yang, Cohen, and Salakhutdinov 2016) exploits label information in a semi-supervised manner. Wang et. al. (2017) propose a matrix factorization model to preserve community structures in network embeddings. However, all these approaches rely on prior knowledge of network structures that are hard to obtain due to technical or privacy issues in many cases.

CDK (Bourigault et al. 2014) offers a clever solution of learning network embeddings from information diffusion

Table 3: Four illustrative communities in Memetracker with typical domains and sample memes.

#1: Technology	#2: Entertainment	#3: Spanish sites	#4: German sites
computerworld.com	accesshollywood.com	madridiario.es	fr-online.de
tech.originalsignal.com	hollywoodreporter.com	abc.es	ad-hoc-news.de
dailytech.com	entertainmentwise.com	informador.com.mx	stern.de
technewsworld.com	celebrity-gossip.net	caracol.com.co	handelsblatt.com
virtual data center os	are you that guy from x-men	atenci n especial	wir sind sehr gl cklich
so what's great about 3g	i'm cool with the paparazzi	la asignatura pendiente	eroberer von davos
solar power plant	luckiest girl in the world (song)	noche de estrellas	alles ist m glich

cascades, but the ranking-based approach does not fully exploit temporal information. Later models that capture temporal dynamics of information contagion (Kurashima et al. 2014; Bourigault, Lamprier, and Gallinari 2016) focus on the local transmission rates that are computationally difficult to infer and prone to over-fitting issues.

To the best of our knowledge, there is no existing work leveraging community structures in social networks in this line of research. However, we consider it important to take the notion of community into consideration: first, community structures prove to be a pervasive phenomenon both empirically and theoretically in network science; second, as mentioned in the introduction, clustering properties in the embedding space prevent over-fitting which is relevant (though, not mathematically strictly equivalent) to low-rank regularization of the user interaction matrix (Zhou, Zha, and Song 2013).

Community Detection. Our work is also inspired by those community detection methods. Newman and Girvan (2002) are among the first to propose that there exist community structures in social networks. After that, community detection becomes a fruitful research area (Papadopoulos et al. 2012) in network science and data mining communities.

Recent work (Kozdoba and Mannor 2015; Yang et al. 2016) attempts to detect communities via learning network embeddings. Barbieri et. al. (2013) first propose to detect communities from diffusion cascades without networks. NetCodec (Long et al. 2015) jointly detects community structure and network infectivity from cascade data.

Information Diffusion. Information diffusion is also a vast research domain, attracting extensive research interests (Guille et al. 2013) with two types of classic information propagation models, Independent Cascade (IC) Model (Kempe, Kleinberg, and Tardos 2003) and Linear Threshold (LT) Model (Granovetter 1978).

Gomez-Rodriguez et. al. propose NETINF (Gomez Rodriguez, Leskovec, and Krause 2010) to infer influence strengths between news media sites and blogs from citation cascades. NetRate (Gomez-Rodriguez, Balduzzi, and Schölkopf 2011) extends the discrete IC model to a continuous IC model and estimates transmission rates between users from cascading data. Later, there are both empirical and theoretical discussions (e.g., (Du et al. 2013; Daneshmand et al. 2014)) on diffusion network recovery from cascades based on the continuous diffusion model, that is also the one we adopt in this paper.

Conclusion

In this paper, we propose a COmmunity-preserving SocIal Network Embedding (COSINE) model. COSINE exploits the information diffusion process to preserve local community structures and reflects rich types of proximity relationships without prior information on network structures. This model can be applied in various tasks such as community detection, information diffusion prediction and visualization. Experimental studies are conducted on both synthetic and real-world datasets, demonstrating its effectiveness, robustness and scalability. Our case study shows that COSINE can help us better understand social contagion and interaction relationships in social networks.

There are various ways to extend our work in the future. One of them is to incorporate semantic information into our social embeddings. We also plan to learn dynamic representations to capture time-varying social networks with evolving community structures.

Acknowledgments

We would thank Zhiqiang Liu for part of the code implementation. We would also like to acknowledge the anonymous reviewers for their valuable comments. This work is supported by 973 Program under Grant No. 2014CB340405, NSFC under Grant No. 61532001 and No.61370054, and MOE-RCOE under Grant No. 2016ZD201.

References

Barbieri, N.; Bonchi, F.; and Manco, G. 2013. Influence-based network-oblivious community detection. In *Proceedings of ICDM'13*, 955–960. IEEE.

Bourigault, S.; Lagnier, C.; Lamprier, S.; Denoyer, L.; and Gallinari, P. 2014. Learning social network embeddings for predicting information diffusion. In *Proceedings of WSDM'14*, 393–402. ACM.

Bourigault, S.; Lamprier, S.; and Gallinari, P. 2016. Representation learning for information diffusion through social networks: an embedded cascade model. In *Proceedings of WSDM'16*, 573–582. ACM.

Cao, S.; Lu, W.; and Xu, Q. 2015. Grarep:learning graph representations with global structural information. In *Proceedings of CIKM'15*, 891–900.

Cox, T. F., and Cox, M. A. 2001. *Multidimensional scaling*. Chapman & Hall/CRC.

Daneshmand, H.; Gomez-Rodriguez, M.; Song, L.; and Schoelkopf, B. 2014. Estimating diffusion network structures: Recovery conditions, sample complexity & soft-thresholding algorithm. In *Proceedings of ICML'14*, volume 32, 793–801. Bejing, China: PMLR.

Du, N.; Song, L.; Gomezrodriguez, M.; and Zha, H. 2013. Scalable influence estimation in continuous-time diffusion networks. In *Proceedings of NIPS'13*, 3147.

Girvan, M., and Newman, M. E. J. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99(12):7821–6.

Gomez-Rodriguez, M.; Balduzzi, D.; and Schölkopf, B. 2011. Uncovering the temporal dynamics of diffusion networks. In *Proceedings of ICML'11*.

Gomez Rodriguez, M.; Leskovec, J.; and Krause, A. 2010. Inferring networks of diffusion and influence. In *Proceedings of KDD'10*, 1019–1028. ACM.

Granovetter, M. 1978. Threshold models of collective behavior. *American Journal of Sociology* 83(6):1420–1443.

Grover, A., and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of KDD'16*. ACM.

Guille, A.; Hacid, H.; Favre, C.; and Zighed, D. A. 2013. Information diffusion in online social networks: A survey. *SIGMOD Record* 42(2):17–28.

Hu, Z.; Yao, J.; Cui, B.; and Xing, E. 2015. Community level diffusion extraction. In *Proceedings of SIGMOD'15*, 1555–1569. ACM.

Karypis, G., and Kumar, V. 1998. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific Computing* 20(1):359–392.

Kempe, D.; Kleinberg, J.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *Proceedings of KDD'03*, 137–146. ACM.

Kozdoba, M., and Mannor, S. 2015. Community detection via measure space embedding. In *Proceedings of NIPS'15*, 2890–2898.

Kurashima, T.; Iwata, T.; Takaya, N.; and Sawada, H. 2014. Probabilistic latent network visualization: inferring and embedding diffusion networks. In *Proceedings of KDD'14*, 1236–1245. ACM.

Lancichinetti, A.; Fortunato, S.; and Radicchi, F. 2008. Benchmark graphs for testing community detection algorithms. *Physical Review. E* 78:4.

Leskovec, J.; Backstrom, L.; and Kleinberg, J. 2009. Meme-tracking and the dynamics of the news cycle. In *Proceedings of KDD'09*, 497–506.

Long, T.; Farajtabar, M.; Song, L.; and Zha, H. 2015. Netcodec: Community detection from individual activities. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, 91–99. SIAM.

Luo, D.; Ding, C. H. Q.; Nie, F.; and Huang, H. 2011. Cauchy graph embedding. In *Proceedings of ICML'11*, 553–560.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of NIPS'13*, 3111–3119.

Papadopoulos, S.; Kompatsiaris, Y.; Vakali, A.; and Spyridonos, P. 2012. Community detection in social media. *Data Mining and Knowledge Discovery* 24(3):515–554.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *Proceedings of KDD'14*, 701–710. ACM.

Roweis, S. T., and Saul, L. K. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290(5500):2323–2326.

Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; and Mei, Q. 2015. Line: Large-scale information network embedding. In *Proceedings of WWW'15*, 1067–1077. ACM.

Teh, Y. W.; Jordan, M. I.; Beal, M. J.; and Blei, D. M. 2006. Hierarchical dirichlet processes. *Journal of the American Statistical Association* 101(476):1566–1581.

Wang, X.; Cui, P.; Wang, J.; Pei, J.; Zhu, W.; and Yang, S. 2017. Community preserving network embedding. In *Proceedings of AAAI'17*.

Whitmore, G. A. 1983. A regression method for censored inverse-gaussian data. *Canadian Journal of Statistics* 11(11):305–315.

Yang, L.; Cao, X.; He, D.; Wang, C.; Wang, X.; and Zhang, W. 2016. Modularity based community detection with deep learning. In *Proceedings of IJCAI'16*.

Yang, Z.; Cohen, W. W.; and Salakhutdinov, R. 2016. Revisiting semi-supervised learning with graph embeddings. In *Proceedings of ICML'16*.

Yang, Z.; Tang, J.; and Cohen, W. W. 2016. Multi-modal bayesian embeddings for learning social knowledge graphs. In *Proceedings of IJCAI'16*, 2287–2293.

Zhou, K.; Zha, H.; and Song, L. 2013. Learning social infectivity in sparse low-rank networks using multi-dimensional hawkes processes. In *Proceedings of AISTATS'13*, PMLR series, 641–649.