Stochastic Opinion Dynamics for Interest Prediction in Social Networks

Diploma Thesis Presentation

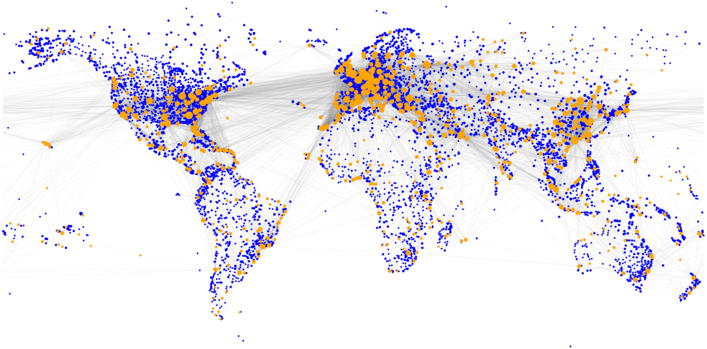
Marios Papachristou

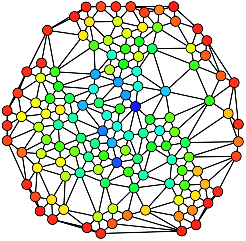
Advisor: Dimitris Fotakis

Thesis Committee: D. Fotakis, A. Pagourtzis, N. Papaspyrou

25th June, 2020

National Technical University of Athens
School of Electrical and Computer Engineering
Division of Computer Science





Motivation i

- Most large-scale Online Social Networks (OSN) exhibit the *core-periphery structure* [14, 31, 23, 32, 26, 22, 29]
  - Nodes are naturally partitioned into
    - a *core set C* of nodes that are tightly connected with each other.
    - a *periphery set U*, where the nodes are sparsely connected, but are relatively well-connected to the core.
  - The theory stems from Wallerstein's **World-systems theory** [29]. Later, Krugman [17] studied the problem from a socio-economical perspective
    - capital-intensive production at the core
    - labour-intensive production at the periphery
    - trade flows and diplomatic ties also follow this structure
  - Mathematical mode
    - Stochastic Blockmodel (under certain assumptions) [31]
    - Continuous Models [14]
    - Discrete Models [4]

Motivation ii

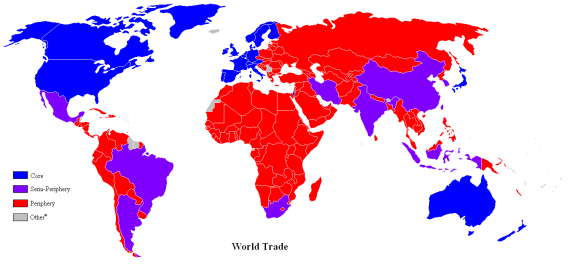


Figure 1: World-Systems Theory as formulated by Wallerstein

- In most of the cases, the a sublinear fraction of the core nodes (e.g. $n^{0.7}$) **almost dominate** the rest of the network, in the sense that a small fraction of δn high-degree nodes dominate an $(1 - a)n$ fraction of the network's engaged nodes (with in-degree above a threshold).

Motivation iii

- Social Networks are highly **homophilic**. People tend to exchange opinions with other people "similar" to them.
- The problem we are studying is important because
  - Profile information of influential nodes is usually public.
  - Leveraging the core-periphery structure of networks is a way to develop very fast algorithms.
  - It has ability to scale to large networks, much faster than other ML methods (e.g. network embeddings).

"Birds of a feather flock together"
— Plato, Symposium, ca. 385 BC

- Identify and use the influencers as **steady-state trend-setters**.
- Inspired by *coevolutionary opinion formation* [13, 1], we next treat the network as the result of a natural *interest exchange* dynamical process.
- Throughout the process, each peripheral user interacts only with her k-nearest neighbors. We call this generative model the **Nearest Neighbor Influence Model (NNIM)**.
- **Novelties**
  - Leverage the sublinear core of social networks to build algorithms that can efficiently scale to millions of nodes.
  - Use homophily as a way to explain the user interactions at the periphery.

Problem Definition

Interest Prediction (a.k.a. Multi-label Classification)

INPUT: A Social Network $G(U, E)$ and a core set $C \subseteq U$ where the core is labeled with binary labels, or interests, $\hat{X}_c \in \{0, 1\}^d$ for all $c \in C$. The core dominates the set U.
OUTPUT: A score vector $\boldsymbol{\phi}_u^* \in [0, 1]^d$ for every $u \in U$ where each entry $\phi_{iu} \in [0, 1]$ corresponds to the probability that user u adopts interest $i \in [d]$.

The domination hypothesis happens to a large fraction of the nodes in real-world datasets.

Real-world Observations: Twitter most followed users (B. Obama, etc.), Instagram influencers (product promoters, activists etc.)

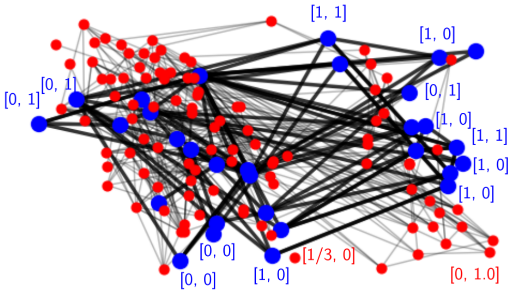


Figure 2: Real-world example (labels are artificial). Blue: Core network. Red: Periphery network.

Related Works

- **Multilabel classification** is a standard benchmark task for graph mining methods, such as node embeddings, usually combined with logistic regression. Examples are: node2vec, DeepWalk, NodeSketch, GraphWave, etc.
- Other similar tasks in graph mining: community detection (BigCLAM [30], ego-circles [19] etc.)
- **Opinion Dynamics** like the Hegselmann-Krausse Model [13], Coevolutionary Opinion Formation [2] etc. (more later)
- **Inference in probabilistic graphical models**
    - Expectation-Maximization [5]
    - Variational Expectation-Maximization
    - Mean-field approximation [15, 27]

We are given a social network $G(U, E)$ associated with a feature vector mapping $\boldsymbol{x} : U \to S \subseteq \mathbb{R}^d$.

The features are in general *socio-demographical*, such as age, gender, political orientation, hobbies etc. A reasonable choice for S can be $[0, 1]^d$ at which for example the i-th component of $\boldsymbol{x}_u$ for some $u \in U$ displays probabilities that user u adopts interest $i \in [d]$.

Homophily Property. For users $u, v \in U$ if $\|\boldsymbol{x}_u - \boldsymbol{x}_v\|$ is small then $\Pr[(u, v) \in E(G)]$ is high.

Homophily ii

How about *neighborhoods* in the network?

We restrict ourselves to features in $[0, 1]^d$. For each user $u \in U$ we form a neighborhood $K(u)$ of her k_u-nearest-neighbors (given the initial features $\hat{X}_v$) and compare it with the actual neighborhood $N(u)$ (in directed networks we use the out-neighborhood $N^+(u)$). We define two vectors

$$\boldsymbol{\alpha}_u = \frac{1}{|N(u)|} \sum_{v \in N(u)} \hat{X}_v \tag{1}$$

$$\boldsymbol{\beta}_u = \frac{1}{|K(u)|} \sum_{v \in K(u)} \hat{X}_v \tag{2}$$

Calculate the weighted average

$$\mathrm{H.I.} = 100\,\% \times \left(1 - \frac{\sum_{u \in U} |N(u)| \mathrm{RMSE}(\boldsymbol{\alpha}_u, \boldsymbol{\beta}_u)}{2|E|}\right) \tag{3}$$

We call this quantity the **Homophilic Index**. The more close $\boldsymbol{\alpha}_u$'s are to $\boldsymbol{\beta}_u$'s the higher the index is.

- A high index means the the k_u-nearest-neighborhoods can almost be used in place of the real neighborhoods.
- **Advantage:** We may end looking at a significantly smaller graph.

Insight into the Data

Table 1: Dataset Statistics and Homophilic Index are reported. We count directed edges where the network is undirected. The Homophilic Index is calculated after dimensionality reduction with PCA so that 95% of the original variance is explained after the transformation.

Name	Network Type	Nodes	Edges	Homophilic Index $\lvert \mathrm{out}(u) \rvert + 1$	(k_u) $\lceil \log n \rceil$	d
facebook [18, 19]	ego	1.03K	27.8K	93.24	91.03	576
dblp-dyn [6]	co-authorship	1.23K	4.6K	82.02	83.56	43
fb-pages [18, 25]	page-page	22.5K	342K	91.69	92.31	4
github [18, 25]	developer	37.7K	578K	85.48	84.41	1
dblp [24]	co-authorship	41.3K	420K	82.54	85.62	29
pokec [18, 28]	social	1.6M	30.6M	66.10	67.72	280

General Idea. Each agent builds a (possibly dynamic) neighborhood and updates her opinion $x_u^{(t)} \in [0,1]^d$ according to the opinions of her and her neighbors.

The Friedkin-Johnsen Model [9]. The adjacency matrix is weighted with weights w_{uv}.

$$x_u^{(t+1)} = \frac{\sum_{v \in N(u)} w_{uv} x_v^{(t)} + w_{uu} x_u^{(0)}}{\sum_{v \in N(u)+u} w_{uv}} \tag{4}$$

The model is stable and converges in $O\left(\frac{\ln(n/\delta)}{1-\rho}\right)$ iterations, where $\rho < 1$ is the spectral radius of the system.

Game-theoretical explaination for the model at Bindel et. al. [3], , and Bhawalkar, Gollapudi and Munagala [1].

Coevolutionary Opinion Formation Games [1]. The setting is modeled as a game of n players where each player has an intrinsic opinion $s_i \in [0,1]$ and expresses an opinion z_i (in general $s_i \neq z_i$) and her goal is to minimize her cost $C_i(\mathbf{z})$ that depends on $\mathbf{z} = (z_1, \ldots, z_n)$. The social cost is $C(\mathbf{z}) = \sum_{i \in [n]} C_i(\mathbf{z})$. They consider two types of games

1. The Symmetric Game, where

$$C_i(z_i, \mathbf{z}_{-i}) = \sum_{j \neq i} f_{ij}(z_i - z_j) + w_i g_i(z_i - s_i)$$

   where f_{ij}, g_i are symmetric (that is $f_{ij}(x) = f_{ij}(-x)$, $g_i(x) = g_i(-x)$ and $g(0) = 0$).

   - PoA of at most 2 for all convex functions f_{ij} and g_i.
   - PoA is tightly bounded for all strictly convex weighting functions. (via Local Smoothness technique)

Opinion Dynamics iv

2. The K-NN Game where each agent forms her k nearest neighbors wrt. her intrinsic opinion s_i and suffers a cost

$$C_i(z_i, \mathbf{z}_{-i}) = \sum_{j \in K(\mathbf{z}, i)} (z_j - z_i)^2 + \alpha k (z_i - s_i)^2$$

   - The game has a PoA of at most a constant, where the constant improves as α increases.
   - Intuitively, the social outcomes become better when nodes are "narrow minded" and give larger weight to their own opinion.
   - The PoA bounded even when agents place almost equal weight to their neighbors.
   - For small α the PoA is at least $1/\alpha^2$, namely it deteriorates as nodes become broad-minded.

The Hegselmann-Krausse Model [13]. Each agent builds the set

$$S_u^{(t)} = \left\{ v \in U \mid \|x_u^{(t)} - x_v^{(t)}\| \leq \epsilon \right\} \tag{5}$$

and updates her opinion according to

$$x_u^{(t+1)} = \frac{1}{|S_u^{(t)}|} \sum_{v \in S_u^{(t)}} x_v^{(t)} \tag{6}$$

Similar models can be found at [8, 7].

However...

- These models *are not generative* $\implies$ Not able to generate data.
- Structural computational barriers (e.g. radius queries for HK, dense graph for FJ)

May be unable to explain phenomena in **large networks** (>100K nodes)

We devise a **stochastic** model for opinion formation driven by homophilic properties of the networks.

Each peripheral agent/user u has an opinion vector $X_u^{(t)} \in \{0,1\}^d$. The **disagreement** between two users is defined as the Hamming Distance Between them

$$\|X_u^{(t)} - X_v^{(t)}\| = \sum_{i=1}^{d} \mathbf{1}\left\{X_{ui}^{(t)} \neq X_{vi}^{(t)}\right\} \tag{7}$$

that is the number of points at which they disagree.

Our Model ii

At each round t the agent looks at his k-nearest neighbors with respect to the Hamming Distance (breaking ties consistently).

We call this set $\mathcal{K}^{(t)}(u)$

Then each user creates the vector $\boldsymbol{\xi}_u^{(t+1)}$ such that

$$\boldsymbol{\xi}_u^{(t+1)} = \frac{1}{k} \sum_{v \in \mathcal{K}^{(t)}(u)} X_v^{(t)} \tag{8}$$

The user updates her opinions as

$$X_u^{(t+1)} \sim \mathrm{Be}\left(\boldsymbol{\xi}_u^{(t+1)}\right) \tag{9}$$

The opinion parameters are **initialized** according to the average of the opinions of the influencers the user is following.

Our Model iii

Given that we only know the initial conditions or the initial probabilities how can we infer what will eventually happen?

Inference in these models is **computationally inefficient** since summation is required in the latent variable space.

Even Bernoulli is exponential, since there are 2^{nd} possible outcomes.

Expectation-Maximization addresses the problem

Many possible variants [5, 15, 27, 12, 16].

At each round each agent “looks” at the instantaneous likelihood

$$\mathcal{L}_{\xi}^{(t+1)}\left(\xi_U^{(t+1)}\right) = \log \sum_{X_U^{(t)}} \Pr\left[X_U^{(t)} | \xi_U^{(t+1)}\right] \tag{10}$$

In general we assume that agents are affected only by the previous step (Markov Property) like in the game-theoretical work of Bindel et al. [3].

We use Jensen’s Inequality to derive a lower bound

$$\mathcal{L}_{\xi}^{(t+1)} \geq \underbrace{\mathbb{E}_{Q^{(t)}}\left[\log \Pr\left[X_U^{(t)} \middle| \xi_U^{(t+1)}\right]\right]}_{\text{ELBO}} + \underbrace{\mathbb{E}_{Q^{(t)}}\left[-\log Q(X_U^{(t)})\right]}_{\text{ENTROPY}} \tag{11}$$

Inference ii

We maximize the ELBO (Evidence Lower Bound) subject to a **variational distribution** $Q^{(t)}$.

The choice of $Q^{(t)}$ can vary (per problem). The most relevant to ours is the **mean-field** method [15, 27], known from Statistical Physics where

$$Q^{(t)} = \prod_{u \in U} \prod_{i=1}^{d} \left(\phi_{iu}^{(t)}\right)^{X_{iu}^{(t)}} \left(1 - \phi_{iu}^{(t)}\right)^{1-X_{iu}^{(t)}} \tag{12}$$

of Independent Bernoulli variables.

Mean-field assumes independence of variables to infer the actual parameters $\boldsymbol{\xi}_u^{(t)}$.

The ELBO bound becomes

$$\mathcal{L}_{Q,\xi}^{(t+1)} = \mathbb{E}_{Q^{(t)}}\left[\sum_{i=1}^{d}\sum_{u\in U}\sum_{v\in S} \mathbb{1}\left\{v \in \mathcal{K}^{(t)}(u)\right\}\left(X_{iv}^{(t)} \log \xi_{iu}^{(t+1)} + \left(1 - X_{iv}^{(t)}\right)\log\left(1 - \xi_{iu}^{(t+1)}\right)\right)\right] \tag{13}$$

We want a way to approximate the sum with the indicator variable inside.

Inference iv

Idea: Concentration Bounds. We show that for two Bernoulli Vectors X, Y the Hamming Distance is close to their parameter vectors (squared Eucledian)

We give the worst case bound

Theorem
Let $X, Y \in \{0,1\}^d$ be two Bernoulli Vectors with parameters $\mathbb{E}[X] = \boldsymbol{p},\ \mathbb{E}[Y] = \boldsymbol{q}$ and pairwise independent components and $\epsilon > 0$ be a positive real number. Then

$$\Pr\left[\left|\|X - Y\| - \|\boldsymbol{p} - \boldsymbol{q}\|\right| > \frac{(1+\epsilon)d}{2}\right] \leq 2\exp\left(-\frac{\epsilon^2 d}{2}\right) \tag{14}$$

Inference v

We use the previous result and prove that it suffices to choose

$$k \leq C\left(4n \exp\left(-\epsilon^2 d\right) + \log n\right) \qquad C > 1 \tag{15}$$

neighbors such that the stochastic set $\mathcal{K}^{(t)}(u)$ approaches the "mean set" $K^{(t)}(u)$ with respect to the parameter space (neighbors are taken according to the distance of their means) a.a.s. for $n \to \infty$.

$$\mathcal{L}_{Q,\xi}^{(t+1)} \approx \sum_{i=1}^{d} \sum_{u \in U} \sum_{v \in K^{(t)}(u)} \left[\phi_{iv}^{(t)} \log \phi_{iu}^{(t+1)} + \left(1 - \phi_{iv}^{(t)}\right) \log\left(1 - \phi_{iu}^{(t+1)}\right)\right] \tag{16}$$

Note that also now $\boldsymbol{\xi}_u^{(t)} \approx \boldsymbol{\phi}_u^{(t)}$ from Hoeffding's inequality.

Inference vi

Taking partial derivatives with respect to $\phi_{iu}^{(t+1)}$

$$\frac{\partial \mathcal{L}_{Q,\xi}^{(t+1)}}{\partial \phi_{iu}^{(t+1)}} = 0 \tag{17}$$

we arrive to the iterative equations

$$\phi_{iu}^{(t+1)} = \frac{1}{k} \sum_{v \in K^{(t)}(u)} \phi_{iv}^{(t)} \tag{18}$$

SIMILAR TO THE DETERMINISTIC OPINION DYNAMICS!

We can also define the Macroscopic Distribution to be determined by

$$\boldsymbol{\mu}^{(t)} = \frac{1}{|U|} \sum_{u \in U} \boldsymbol{\xi}_u^{(t)} \tag{19}$$

Using similar arguments

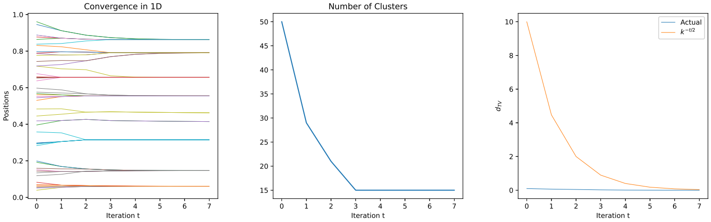
$$\boldsymbol{\mu}^{(t)} = \frac{1}{|U|} \sum_{v \in U} \boldsymbol{\phi}_v^{(t)} \tag{20}$$

Relation to EM

E-STEP: Calculate k-nearest neighbors

M-STEP: Calculate $\boldsymbol{\phi}_u^{(t+1)}$ and $\boldsymbol{\mu}^{(t+1)}$ and repeat until convergence.

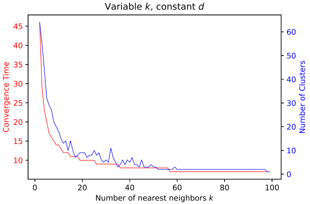
Example (in 1D) i



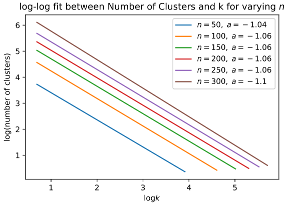
(a) Agent positions (b) Number of clusters (c) Convergence Rate

Figure 3: Microscopic properties of the NNIM model for $n = 100$ agents, $D = 10^{-3}$, and $k = 3$ neighbors.

Example (in 1D) ii



(a) Variable k, constant d



(b) log-log plot between number of clusters and k

Figure 4: Macroscopic Properties of the NNIM model. We have run the with $D = 10^{-7}$.

Convergence i

We use Lyapunov Stability Theory to prove finite time convergence (under breaking ties in a consistent fashion). We write the equations in the form

$$\Phi(t+1) = A(t)\Phi(t) \tag{21}$$

where $A(t)$ has entries $1/k$ in the places where i is in the kNN of j and is zero everywhere else.

This sequence admits an adjoint sequence $\Pi(t)$ with elements $\pi_u^{(t)} > p$ for some $p \in (0,1)$. The adjoint sequence obeys

$$\Pi^T(t+1) = \Pi^T(t)A(t) \tag{22}$$

Convergence ii

We define the potential

$$V(t) = \sum_{i=1}^{n} \pi_u(t) \|\phi_u(t) - \Pi^T(t)\Phi(t)\|_2^2 \tag{23}$$

Which is equivalent to

$$V(t) = V(t+1) + \frac{1}{2} \sum_{u,v} H_{uv}(t)(\phi_u^{(t)} - \phi_v^{(t)})^2 \tag{24}$$

$$H(t) = A^T(t)\mathrm{diag}(\pi_u(t+1))A(t) \tag{25}$$

$$H_{uv}(t) = \frac{1}{k^2} \sum_{w} \pi_w(t+1)\mathbf{1}\{u \in K^{(t)}(w)\}\mathbf{1}\{v \in K^{(t)}(w)\} \tag{26}$$

Hence (except from the equilibrium) the function is negative definite (away from the equilibrium point)

$$V(t+1) = V(t) - \underbrace{\frac{1}{2k^2}\sum_{w}\pi_w(t+1)\sum_{u,v\in K^{(t)}(w)}(\phi_u^{(t)} - \phi_v^{(t)})^2}_{>0} < V(t) \quad (27)$$

Hence system is **Globally Asymptotically Stable**. For $t \to \infty$ the agents form clusters $\sigma^{(t)}(u)$ for each $t \geq 0$ and $u \in U$.

Convergence iv

For two contiguous sets of agents W, Z we define the metric

$$\delta_{WZ}^{(t)} = \min_{w \in W, z \in Z} \|\phi_w^{(t)} - \phi_z^{(t)}\| \tag{28}$$

Two sets isolate if the distance δ_{WZ} becomes "large enough" that anyone of W's neighbors does not lie in Z and vice versa.

Observation: If two sets isolate at some t_0 they remain isolated for $t \geq t_0$.

FINITE TIME CONVERGENCE $\implies$ GOOD ALGORITHM

Convergence v

For deriving D-accuracy bounds we consider the **mixing time** of the process (identical to a Markov Chain).

Convergence depends on the second largest eigenvalue of the slowest matrix

$$\lambda_2^* = \max_{1 \leq t \leq T} \{\lambda_2(A(t))\} \tag{29}$$

All matrices correspond to k-regular graphs! Well-known and non-trivial problem (unsolved for many years) conjectured by Alon and proved by Friedman [10]

We need $T \geq \lceil 2 \log(1/D) \log^{-1} k \rceil$
iterations to be $d \cdot D$-close to convergence
(in terms of total variation).

Table 2: Complexity of NNIM with under various data structures (Brute-force, KD-tree, Metric Ball, LSH) for running the NNIM model such that the total variation distance is at most $D > 0$ after execution. State-of-the-art is DCI and Prioritized DCI [20, 21]. The quantity d' is the intrinsic dimension, such that for a dataset any ball of radius r contains $O(r^{d'})$ points. The quantity m refers to the number of projection directions used in the DCI/Prioritized DCI algorithm.

Data structure	Complexity	Notes
Brute-force	$O\left(nd(n+k)\log(1/D)\log^{-1}k\right)$	Efficient for very small n
KD/Ball tree	$O\left(nd(n^{1-1/d}+k)\log(1/D)\log^{-1}k\right)$	Efficient for $d \ll \log n$
LSH [11]	$O\left(n^{1+1/(1+\epsilon)^2}dk\log(1/D)\log^{-1}k\right)$	$(1+\epsilon)$-approximation
DCI/PDCI [20, 21]	$O\left(\left(mn\log\left(\frac{n}{k}\right)+\left(\frac{n}{k}\right)^{2-\frac{m}{d'}}\right)\frac{dk\log(1/D)}{\log k}\right)$	Efficient for large n and d'

Scaling

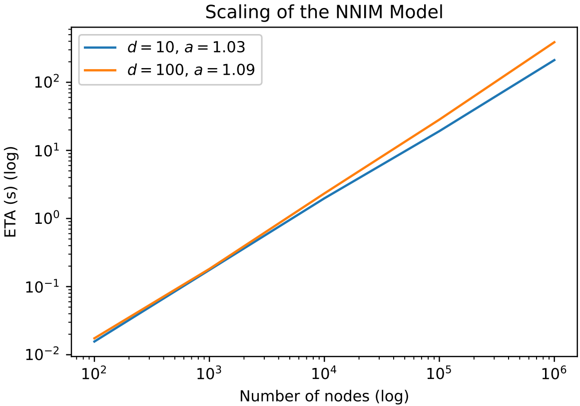


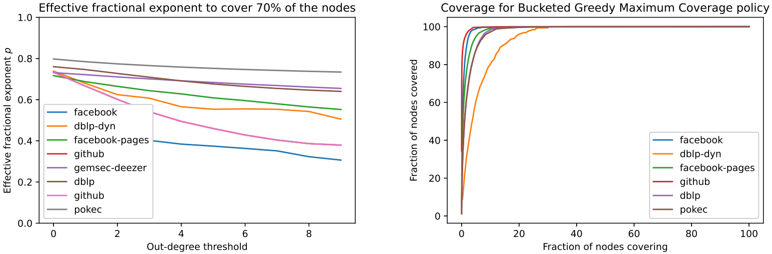
Figure 5: Log-log plot of the total time taken to perform inference to a network of up to 1M agents and $d \in \{10, 100\}$ with binary equiprobable artificial features, $D = 0.001$ and $k = \lceil \log n \rceil$. using LSH to obtain the nearest neighbors.

Experiments

We infer the hobbies of users (make recommendations) without observing the whole underlying network. We identify the K core users as follows

1. We sort the nodes according to their in-degree and put them into $\log(n/K)/\log\gamma$ non-uniform buckets $V_1, \ldots, V_r, \ldots$ of sizes $\lceil\gamma K\rceil, \ldots, \lceil\gamma^r K\rceil - \lceil\gamma^{r-1} K\rceil, \ldots$, for some $\gamma > 1$.
2. We start by constraining the neighborhoods of vertices to V_1 and run the greedy maximum coverage algorithm on it (remove the nodes with most uncovered neighbors and repeat).
3. Continue to next sets until we exhaust K.
4. The algorithm has an approximation ratio of $1 - 1/e - o(1)$ for the Maximum Coverage Problem.
5. In practice it is very fast! And it generates very good results: for an in-degree threshold of $\tau = 4$ a population of $n^{0.7}$ influencers dominate $> 70\%$ of the network.

We compare our approach with network embeddings (node2vec, GraphWave, NodeSketch) and the Random-HK model.



(a) (b)

Figure 6: Left: Engagement Threshold Effect. Right: Coverage curve for the BGMC policy for $\tau = 4$.

Results i

	facebook	dblp-dyn	fb-pages	github	dblp	pokec
			AUC-ROC (all labels)			
node2vec	**86.35**	87.42	84.00	67.23	69.80	**96.93**
GraphWave	86.20	86.78	70.96	45.13	69.57	†
NodeSketch	80.90	81.90	68.68	49.96	58.88	92.14
Random HK ($k = \lceil \log n \rceil$)	84.75	86.30	71.90	50.34	68.83	†
NNIM ($k = \lceil \log n \rceil$)	84.24	88.05	**91.86**	68.07	78.64	85.60
NNIM ($k = \lceil \sqrt{n} \rceil$)	85.82	**91.16**	91.62	67.86	**81.65**	91.84
NNIM w/ Reg ($k = \lceil \log n \rceil$)	84.17	87.39	91.78	**72.31**	78.86	85.05
			RMSE (all labels)			
node2vec	0.012	0.059	0.093	0.438	0.166	**0.022**
GraphWave	**0.010**	0.052	7e-6	0.400	**0.082**	†
NodeSketch	0.096	0.123	0.098	0.440	0.316	0.128
Random HK ($k = \lceil \log n \rceil$)	**0.010**	0.056	**4e-17**	0.412	0.096	†
NNIM ($k = \lceil \log n \rceil$)	0.011	0.062	**4e-17**	0.389	0.143	0.026
NNIM ($k = \lceil \sqrt{n} \rceil$)	**0.010**	**0.050**	**4e-17**	**0.388**	0.128	0.025
NNIM w/ Reg ($k = \lceil \log n \rceil$)	0.012	0.066	4e-16	**0.388**	0.145	0.025

Results ii

	facebook	dblp-dyn	fb-pages	github	dblp	pokec
	AUC-ROC (top 50% of labels)					
node2vec	54.98	**94.92**	78.69	67.23	68.53	**96.94**
GraphWave	53.97	92.91	40.11	45.13	65.70	†
NodeSketch	55.91	92.37	46.50	49.96	58.13	92.14
Random HK ($k = \lceil \log n \rceil$)	52.82	93.10	56.14	50.34	64.49	†
NNIM ($k = \lceil \log n \rceil$)	59.08	79.32	**89.00**	68.27	78.69	85.80
NNIM ($k = \lceil \sqrt{n} \rceil$)	58.30	90.59	88.04	67.86	**80.85**	91.84
NNIM w/ Reg ($k = \lceil \log n \rceil$)	**59.20**	81.11	88.65	**72.31**	79.10	85.05
	AUC-ROC (top-1 label)					
node2vec	52.56	62.82	80.17	67.23	60.28	**55.87**
GraphWave	**57.19**	67.00	61.37	45.13	52.89	†
NodeSketch	53.02	63.06	59.07	49.96	49.22	50.78
Random HK ($k = \lceil \log n \rceil$)	50.17	48.40	49.48	50.34	49.96	†
NNIM ($k = \lceil \log n \rceil$)	53.29	82.89	90.18	68.27	70.31	54.64
NNIM ($k = \lceil \sqrt{n} \rceil$)	53.62	**84.16**	**90.38**	67.86	**71.27**	55.34
NNIM w/ Reg ($k = \lceil \log n \rceil$)	51.52	80.47	90.35	**72.31**	70.71	54.59
Coverage (%)	88.36	97.16	72.20	68.61	66.04	51.70
Influencers (Core size) (%)	12.47	11.83	4.94	4.23	4.12	1.92

Results iii

Method	Runtime (s) for pokec experiment
node2vec	$\sim 10^3$
GraphWave	†
NodeSketch	$\sim 10^3$
Random HK ($k = \lceil \log n \rceil$)	†
NNIM ($k = \lceil \log n \rceil$)	$\sim 10^1$
NNIM ($k = \lceil \sqrt{n} \rceil$)	$\sim 10^2$
NNIM ($k = \lceil \log n \rceil$) with Reg.	$\sim 10^1$

Table 3: Runs 100-times faster in the pokec network (order of 1M nodes)

Conclusions

- Leverage characteristics of networks to develop fast algorithms
    - A small core allows for efficient scaling of algorithmic tasks
    - Homophilic properties "catch" an underlying structure of the network and can account for user interactions, missing links etc.
- Derive a generative model for stochastic opinion exchange
- Derive algorithms for inference using (Variational) EM
- Establish a link between our model and related work
- Prove convergence and convergence bounds
- Perform experiments in very large networks
- Results submitted to NeurIPS 2020.

THANK YOU!

QUESTIONS?

References i

Kshipra Bhawalkar, Sreenivas Gollapudi, and Kamesh Munagala.
Coevolutionary opinion formation games.
In *Proc. of the 45th ACM Symposium on Theory of Computing Conference (STOC 2013)*, pages 41–50. ACM, 2013.

Kshipra Bhawalkar, Sreenivas Gollapudi, and Kamesh Munagala.
Coevolutionary opinion formation games.
In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 41–50, 2013.

David Bindel, Jon Kleinberg, and Sigal Oren.
How bad is forming your own opinion?
Games and Economic Behavior, 92:248–265, 2015.

Stephen P Borgatti and Martin G Everett.
Models of core/periphery structures.
Social networks, 21(4):375–395, 2000.

Arthur P Dempster, Nan M Laird, and Donald B Rubin.
Maximum likelihood from incomplete data via the em algorithm.
Journal of the Royal Statistical Society: Series B (Methodological), 39(1):1–22, 1977.

Elise Desmier, Marc Plantevit, Céline Robardet, and Jean-François Boulicaut.
Cohesive co-evolution patterns in dynamic attributed graphs.
In *International Conference on Discovery Science*, pages 110–124. Springer, 2012.

Dimitris Fotakis, Vardis Kandiros, Vasilis Kontonis, and Stratis Skoulakis.
Opinion dynamics with limited information.
In *International Conference on Web and Internet Economics*, pages 282–296. Springer, 2018.

Dimitris Fotakis, Dimitris Palyvos-Giannas, and Stratis Skoulakis.
Opinion dynamics with local interactions.
In *IJCAI*, pages 279–285, 2016.

Noah E Friedkin and Eugene C Johnsen.
Social influence and opinions.
Journal of Mathematical Sociology, 15(3-4):193–206, 1990.

Joel Friedman.
A proof of Alon's second eigenvalue conjecture and related problems.
American Mathematical Soc., 2008.

Aristides Gionis, Piotr Indyk, Rajeev Motwani, et al.
Similarity search in high dimensions via hashing.
In *Vldb*, volume 99, pages 518–529, 1999.

Gail Gong and Francisco J Samaniego.
Pseudo maximum likelihood estimation: theory and applications.
The Annals of Statistics, pages 861–869, 1981.

Rainer Hegselmann, Ulrich Krause, et al.
Opinion dynamics and bounded confidence models, analysis, and simulation.
Journal of artificial societies and social simulation, 5(3), 2002.

Junteng Jia and Austin R Benson.
Random spatial network models for core-periphery structure.
In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 366–374. ACM, 2019.

Leo P Kadanoff.
More is the same; phase transitions and mean field theories.
Journal of Statistical Physics, 137(5-6):777, 2009.

Myunghwan Kim and Jure Leskovec.
Multiplicative attribute graph model of real-world networks.
Internet mathematics, 8(1-2):113–160, 2012.

Paul Krugman.
Increasing returns and economic geography.
Journal of political economy, 99(3):483–499, 1991.

Jure Leskovec and Andrej Krevl.
SNAP Datasets: Stanford large network dataset collection.
`http://snap.stanford.edu/data`, June 2014.

Jure Leskovec and Julian J Mcauley.
Learning to discover social circles in ego networks.
In *Advances in neural information processing systems*, pages 539–547, 2012.

Ke Li and Jitendra Malik.
Fast k-nearest neighbour search via dynamic continuous indexing.
In *International Conference on Machine Learning*, pages 671–679, 2016.

Ke Li and Jitendra Malik.
Fast k-nearest neighbour search via prioritized dci.
In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 2081–2090. JMLR. org, 2017.

Roger J Nemeth and David A Smith.
International trade and world-system structure: A multiple network analysis.
Review (Fernand Braudel Center), 8(4):517–560, 1985.

Vilfredo Pareto.
The mind and society, volume 1.
Harcourt, Brace and Howe, 1935.

Adriana Prado, Marc Plantevit, Céline Robardet, and Jean-Francois Boulicaut.
Mining graph topological patterns: Finding covariations among vertex descriptors.
IEEE Transactions on Knowledge and Data Engineering, 25(9):2090–2104, 2012.

Benedek Rozemberczki, Carl Allen, and Rik Sarkar.
Multi-scale attributed node embedding, 2019.

David Snyder and Edward L Kick.
Structural position in the world system and economic growth, 1955-1970: A multiple-network analysis of transnational interactions.
American journal of Sociology, 84(5):1096–1126, 1979.

HE Stanley.
Mean field theory of magnetic phase transitions.
Introduction to Phase Transitions and Critical Phenomena, 1971.

Lubos Takac and Michal Zabovsky.
Data analysis in public social networks.
In *International scientific conference and international workshop present day trends of innovations*, volume 1, 2012.

Immanuel Wallerstein.
World-systems analysis, 1987.

References x

Jaewon Yang and Jure Leskovec.
Overlapping community detection at scale: a nonnegative matrix factorization approach.
In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596, 2013.

Xiao Zhang, Travis Martin, and Mark EJ Newman.
Identification of core-periphery structure in networks.
Physical Review E, 91(3):032803, 2015.

Xiao Zhang, Travis Martin, and Mark EJ Newman.
Identification of core-periphery structure in networks.
Physical Review E, 91(3):032803, 2015.