Learning Node Representations for Information-rich Graphs

Major Area Examination

Zexi Huang

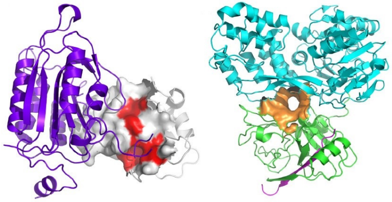
Committee: Ambuj Singh (Chair), Yu-Xiang Wang, Noah Friedkin

Department of Computer Science, University of California, Santa Barbara

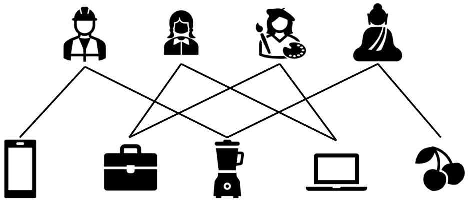
May 28, 2021

Graphs are ubiquitous

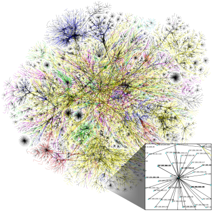
(a) Social network



(b) Protein-protein interactions



(c) Recommendation system



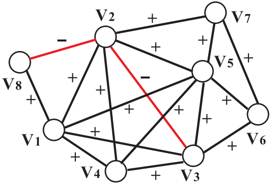
(d) the Internet

Information-rich graphs

(a) Multiscale graph

Information-rich graphs

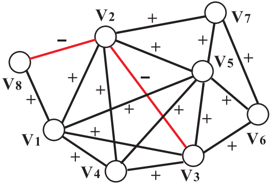
(a) Multiscale graph



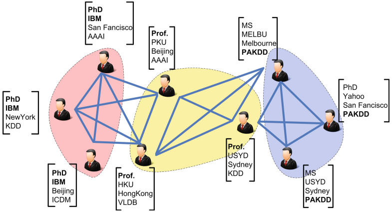
(b) Signed graph

Information-rich graphs

(a) Multiscale graph



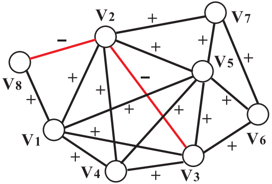
(b) Signed graph



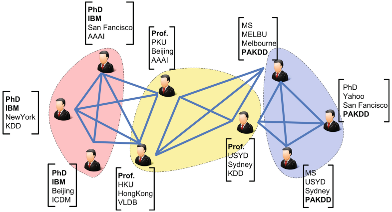
(c) Attributed graph

Information-rich graphs

(a) Multiscale graph



(b) Signed graph



(c) Attributed graph

(d) Physical graph

Social polarization

Donald J. Trump
@realDonaldTrump
Following

TODAY WE MAKE AMERICA GREAT AGAIN!

Donald J. Trump
@realDonaldTrump
Follow

The Fake News is working overtime. Just reported that, despite the tremendous success we are having with the economy & all things else, 91% of the Network News about me is negative (Fake). Why do we work so hard in working with the media when it is corrupt? Take away credentials?

Donald J. Trump
@realDonaldTrump

These are the things and events that happen when a sacred landslide election victory is so unceremoniously & viciously stripped away from great patriots who have been badly & unfairly treated for so long. Go home with love & in peace. Remember this day forever!

This claim of election fraud is disputed, and this Tweet can't be replied to, Retweeted, or liked due to a risk of violence

@

Replying to @realDonaldTrump

These damned fake news outlets need to be taken down! If they had done this to Obama, it would've been the end of the world!

We are with you, Mr President!

@

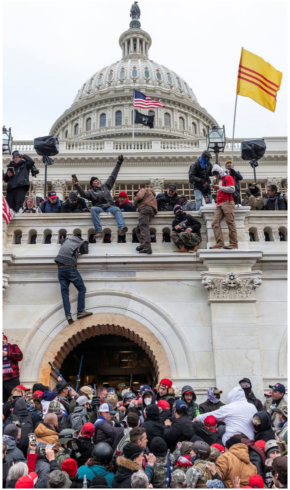
Replying to @realDonaldTrump

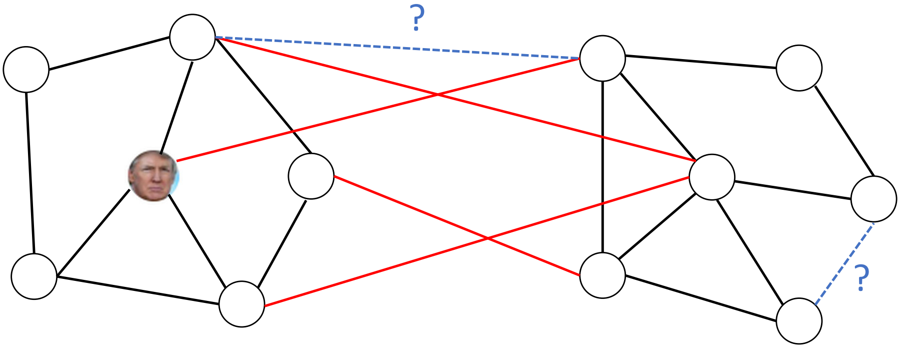
The [D] party will cease to exist once it's all exposed.
FAKE NEWS can no longer control [dampen] public awareness of the TRUTH.
DARK TO LIGHT.
Q

@

Replying to @realDonaldTrump

I love seeing Trump supporters CRY, it's my daily medicine, my weekly energy, my monthly inspiration and my yearly motivation. Their loss is the only reason i'm still alive, i was born to love and enjoy the failure that they have achieved.





Quantifying polarization & signed link prediction

Graph-based applications

- (Signed) link prediction
- Node classification
- Community detection
- Graph regression

Representation learning: extract useful information from data

… for images

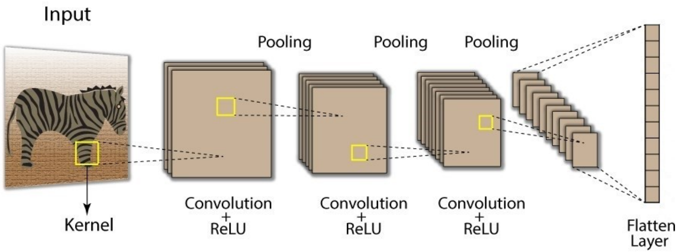
- e.g., convolutional neural networks (CNN)

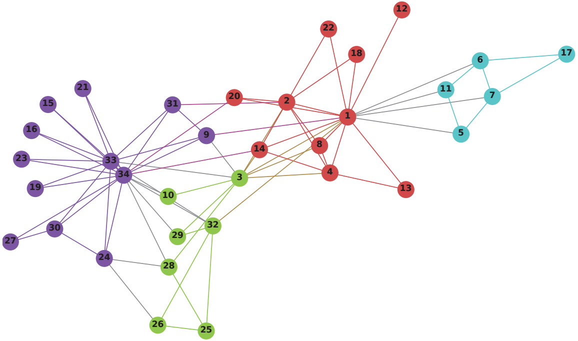
… for text

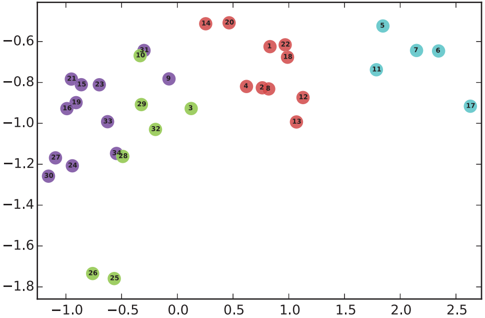
- e.g., skip-gram, transformers

… for graphs

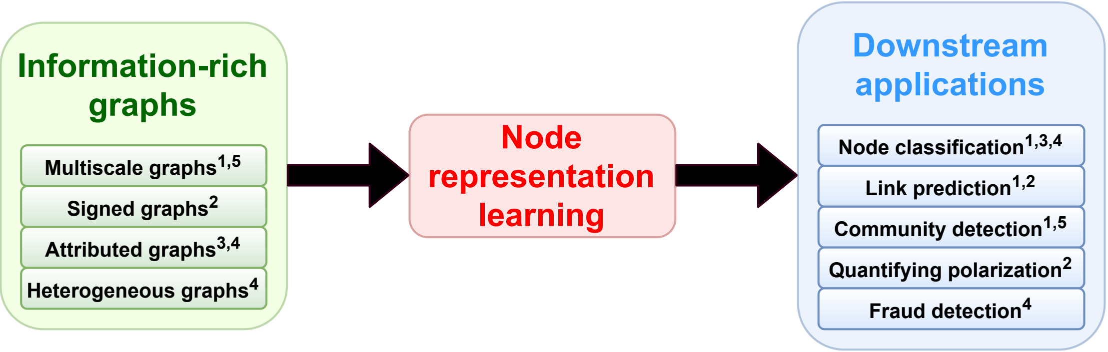
International Conference on Learning Representations (ICLR)







Nodes in a graph $\xrightarrow{\textbf{representation learning}}$ **feature vectors (embeddings)**



[1] **Huang**, Silva, Singh. A broader picture of random-walk based graph embedding. KDD'21.
[2] **Huang**, Silva, Singh. Signed embedding for polarized graphs. Working draft.
[3] Ye, **Huang**, Singh. Krylov graph convolutional networks. Under review.
[4] **Huang**. Graph-based fraud detection in Kindle Direct Publishing. Amazon internship report.
[5] **Huang***, Kondapaneni*, Silva, Singh. Multiscale community detection based on PMI. Ongoing.

Outline

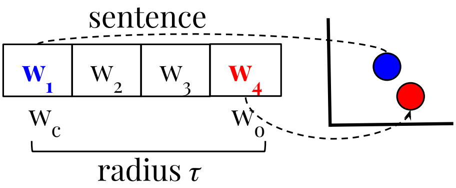
Earlier work

- Spectral graph theory[6]
- Nonlinear dimensionality reduction[7]
- Graph drawing[8]

Deep learning

- Skip-gram for text: word2vec[9]
- Skip-gram for graph: DeepWalk[10]

[6] Chung. Spectral graph theory. 1997.
[7] Belkin, Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. NeurIPS'02.
[8] Díaz, Petit, Serna. A survey of graph layout problems. CSUR'02.
[9] Mikolov et al. Distributed representations of words and phrases and their compositionality. NeurIPS'13.
[10] Perozzi, Al-Rfou, Skiena. Deepwalk: Online learning of social representations. KDD'14.



Earlier work

- Spectral graph theory[6]
- Nonlinear dimensionality reduction[7]
- Graph drawing[8]

Deep learning

- Skip-gram for text: word2vec[9]
- Skip-gram for graph: DeepWalk[10]

---

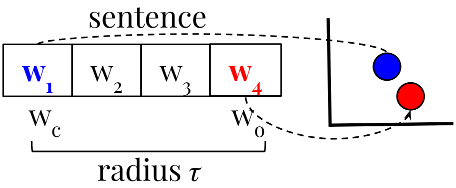
[6] Chung. Spectral graph theory. 1997.
[7] Belkin, Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. NeurIPS'02.
[8] Díaz, Petit, Serna. A survey of graph layout problems. CSUR'02.
[9] Mikolov et al. Distributed representations of words and phrases and their compositionality. NeurIPS'13.
[10] Perozzi, Al-Rfou, Skiena. Deepwalk: Online learning of social representations. KDD'14.

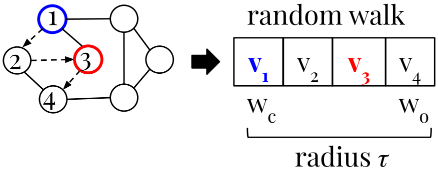
Earlier work

- Spectral graph theory[6]
- Nonlinear dimensionality reduction[7]
- Graph drawing[8]

Deep learning

- Skip-gram for text: word2vec[9]
- Skip-gram for graph: DeepWalk[10]





[6] Chung. Spectral graph theory. 1997.
[7] Belkin, Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. NeurIPS'02.
[8] Díaz, Petit, Serna. A survey of graph layout problems. CSUR'02.
[9] Mikolov et al. Distributed representations of words and phrases and their compositionality. NeurIPS'13.
[10] Perozzi, Al-Rfou, Skiena. Deepwalk: Online learning of social representations. KDD'14.

Different random-walks

- node2vec[11]: biased towards BFS/DFS
- APP[12]: rooted PageRank

Different embedding algorithms

- NefMF[13]: explicit matrix factorization
- NetSMF[14]: spectral sparsifiers for efficiency

Different similarity metrics

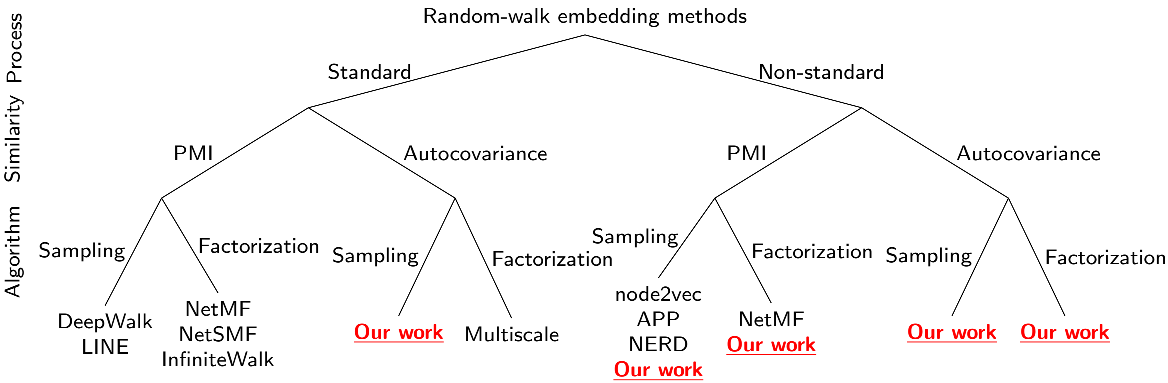
- Stability[15]: autocovariance for multiscale community detection
- Multiscale[16]: autocovariance embedding as a special case

[11] Grover, Leskovec. node2vec: Scalable feature learning fornetworks. KDD'16.
[12] Zhou et al. Scalable graph embedding for asymmetric proximity. AAAI'17.
[13] Qiu et al. Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. WSDM'18.
[14] Qiu et al. Netsmf: Large-scale network embedding as sparse matrix factorization. WebConf'19.
[15] Delvenne et al. Stability of graphcommunities across time scales. PNAS'10.
[16] Schaub et al. Multiscale dynamical embeddings of complex networks. PRE'19.

Problems & questions:

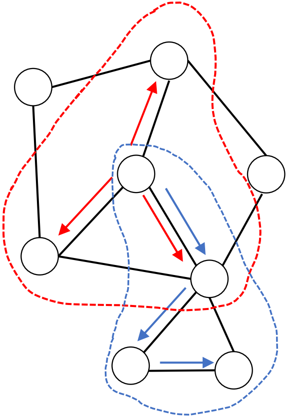
- Difficult to compare existing methods and to advance the SOTA.
- How should embeddings be used for link prediction?
- How do embeddings capture different structural scales?

A broader picture[1]



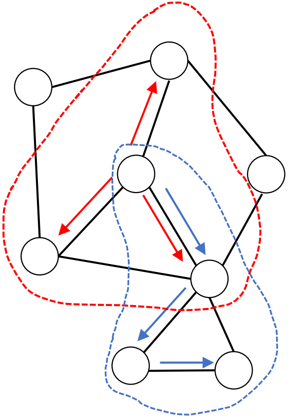
[1] **Huang**, Silva, Singh. A broader picture of random-walk based graph embedding. KDD'21.

Our framework

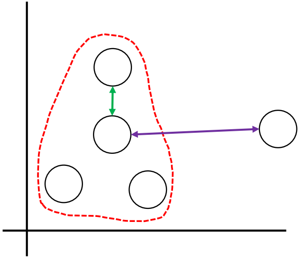


Random-walk process

Our framework

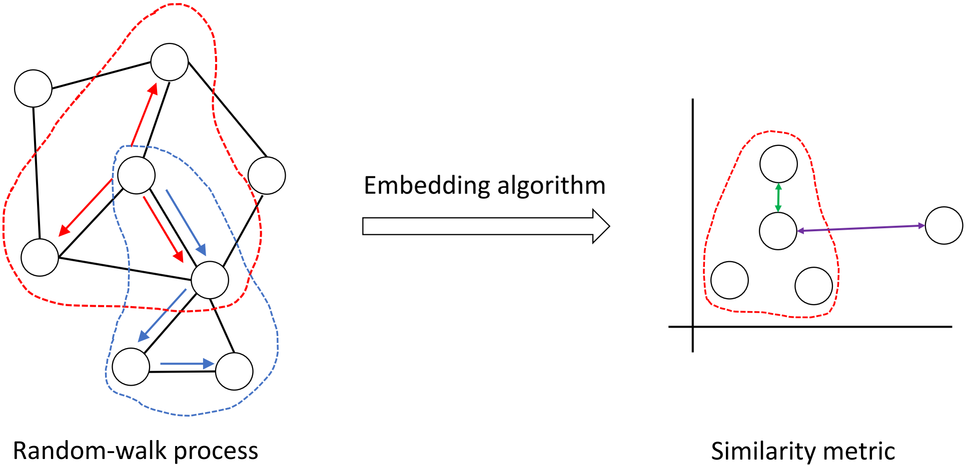


Random-walk process



Similarity metric

Our framework



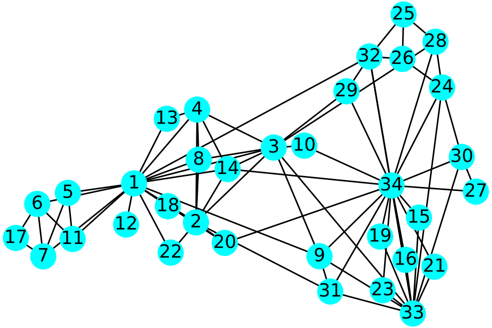
Comparing similarity metrics

PMI: $R = \log(\Pi M^{\tau}) - \log(\pi \pi^{T})$

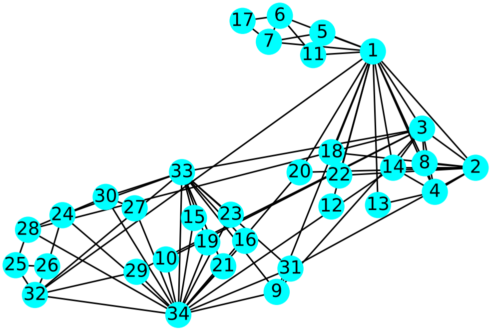
Autocovariance: $R = \Pi M^{\tau} - \pi \pi^{T}$

Comparing similarity metrics

PMI: $R = \color{red}{\log}(\Pi M^\tau) - \color{red}{\log}(\pi \pi^T)$



Autocovariance: $R = \Pi M^\tau - \pi \pi^T$



Comparing similarity metrics

PMI: $R = \log(\Pi M^\tau) - \log(\pi\pi^T)$ Autocovariance: $R = \Pi M^\tau - \pi\pi^T$

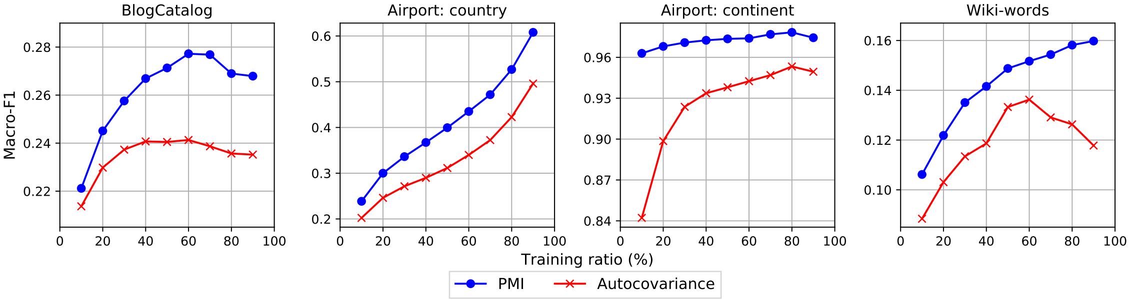


Figure: PMI consistently outperforms autocovariance in node classification.

Comparing similarity metrics

PMI: $R = \color{red}{\log}(\Pi M^{\tau}) - \color{red}{\log}(\pi\pi^{T})$ Autocovariance: $R = \Pi M^{\tau} - \pi\pi^{T}$

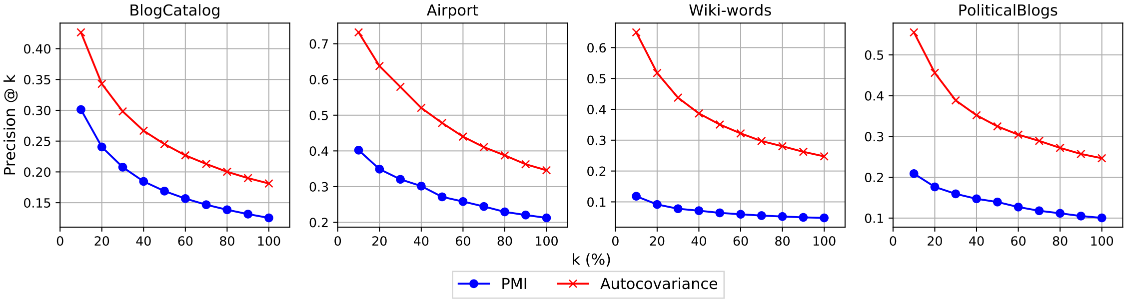


Figure: Autocovariance consistently outperforms PMI in link prediction.

Comparing similarity metrics

$$\text{predicted degree} \propto \text{embedding norm} \propto \begin{cases} \text{actual degree} & \text{for autocov.} \\ \text{constant} & \text{for PMI} \end{cases}$$

Autocovariance captures heterogeneous degree distribution in graphs!

Comparing similarity metrics

$$\text{predicted degree} \propto \text{embedding norm} \propto \begin{cases} \text{actual degree} & \text{for autocov.} \\ \text{constant} & \text{for PMI} \end{cases}$$

Autocovariance captures heterogeneous degree distribution in graphs!

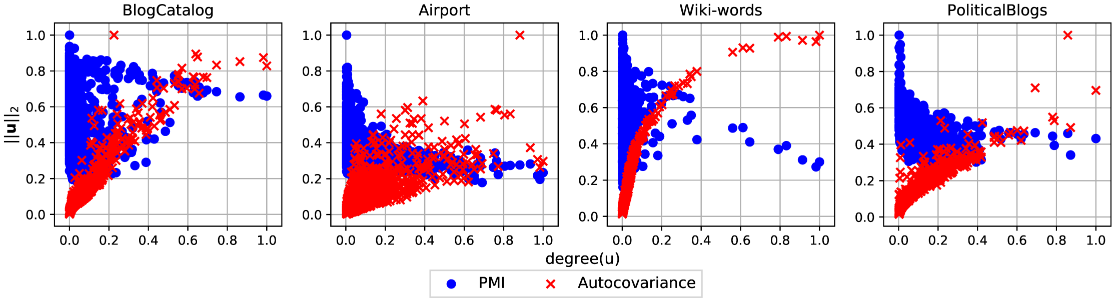


Figure: Autocovariance embedding norms correlated with actual degrees, but not PMI.

Multiscale

PMI: $R = \log(\Pi M^{\tau}) - \log(\pi\pi^T)$ Autocovariance: $R = \Pi M^{\tau} - \pi\pi^T$

Multiscale

PMI: $R = \log(\Pi M^{\tau}) - \log(\pi\pi^T)$

Autocovariance: $R = \Pi M^{\tau} - \pi\pi^T$

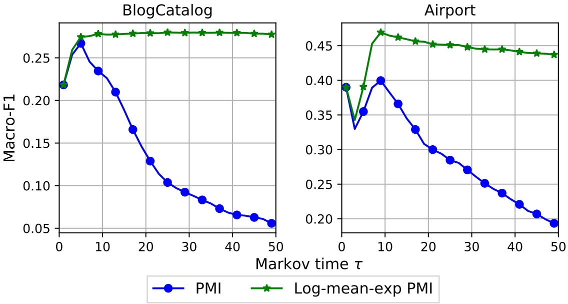


Figure: Node classification performance for PMI can be improved by smooth-averaging across multiple Markov times.

Multiscale

PMI: $R = \log(\Pi M^{\tau}) - \log(\pi\pi^T)$

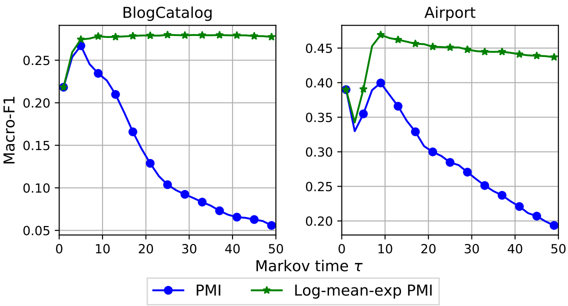


Figure: Node classification performance for PMI can be improved by smooth-averaging across multiple Markov times.

Autocovariance: $R = \Pi M^{\tau} - \pi\pi^T$

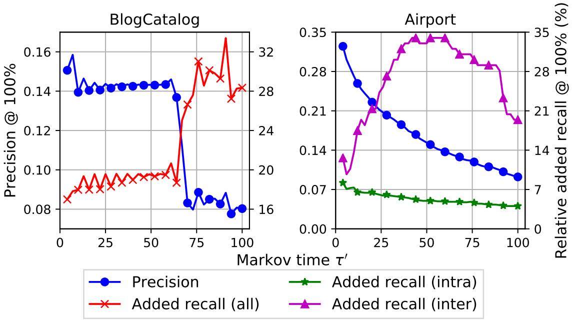


Figure: Prediction of edges of specific structural scales can be improved with different Markov times for autocovariance.

Problems & questions:

- Difficult to compare existing methods and to advance the SOTA.
- How should embeddings be used for link prediction?
- How do embeddings capture different structural scales?

Contributions:

- A unified view of different processes, similarities, and algorithms.
- Autocovariance embedding is significantly better for link prediction.
- Ways to exploit multiscale similarity for optimized performance.

PMI-based multiscale community detection[5]

- Stability[15,17]: multiscale community detection based on clustered autocovariance
- + PMI similarity
- with Louvain algorithm[18].

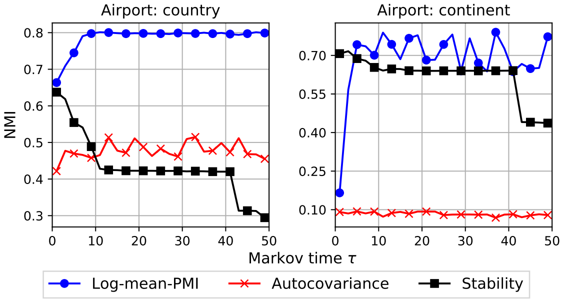


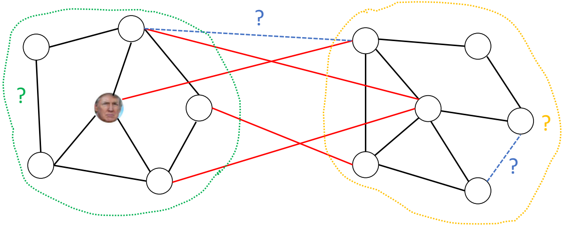
Figure: PMI outperforms both methods based on autocovariance in community detection.

[5] **Huang***, Kondapaneni*, Silva, Singh. Multiscale community detection based on PMI. Ongoing.
[15] Delvenne et al. Stability of graphcommunities across time scales. PNAS'10.
[17] Patelli et al. Generalized Markov stability of network communities. PRE'20.
[18] Blondel et al. Fast unfolding of communities in large networks. JSTAT'08.

Outline

Signed graphs: friendly (+) and adversarial (−) relationships

- e.g., bitcoin markets[19], U.S. Congress[20], online social networks[21]
- Signed link prediction[22]
- Polarized community detection[23]



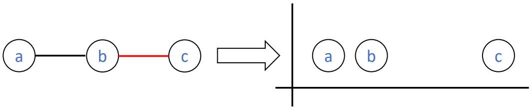
[19] Kumar et al. Edge weight prediction in weighted signed networks. ICDM'16.

[20] Thomas, Pang, Lee. Get out the vote: determining support or opposition from congressional floor-debate transcripts. EMNLP'06.

[21] Lai et al. Stance evolution and Twitter interactions in an Italian political debate. NLDB'18.

[22] Beigi et al. Signed link prediction with sparse data: The role of personality information. WebConf'19.

[23] Bonchi et al. Discovering polarized communities in signed networks. CIKM'19.



Social balance theory[24]:

- SiNE[25]
- SIGNet[26]
- SIDE[27]

Social status theory[28]:

- BESIDE[29]

Network transformation[30]

---

[24] Heider. Attitudes and cognitive organization. J. Psychol.'46.
[25] Wang et al. Signed network embedding in social media. SDM'17.
[26] Islam et al. Signet: Scalable embeddings for signed networks. PAKDD'18.
[27] Kim et al. Side: representation learning in signed directed networks. WebConf'18.
[28] Guha et al. Propagation of trust and distrust. WebConf'04.
[29] Chen et al. "Bridge" enhanced signed directed network embedding. CIKM'18.
[30] Javari et al. ROSE: Role-based signed network embedding. WebConf'20.

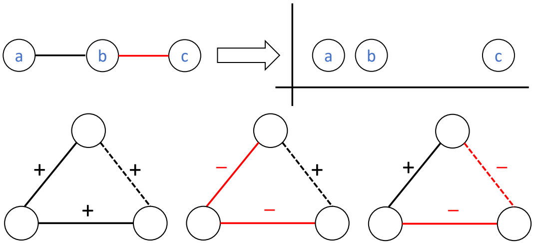
Social balance theory[24]:

- SiNE[25]
- SIGNet[26]
- SIDE[27]

Social status theory[28]:

- BESIDE[29]

Network transformation[30]



[24] Heider. Attitudes and cognitive organization. J. Psychol.'46.
[25] Wang et al. Signed network embedding in social media. SDM'17.
[26] Islam et al. Signet: Scalable embeddings for signed networks. PAKDD'18.
[27] Kim et al. Side: representation learning in signed directed networks. WebConf'18.
[28] Guha et al. Propagation of trust and distrust. WebConf'04.
[29] Chen et al. "Bridge" enhanced signed directed network embedding. CIKM'18.
[30] Javari et al. ROSE: Role-based signed network embedding. WebConf'20.

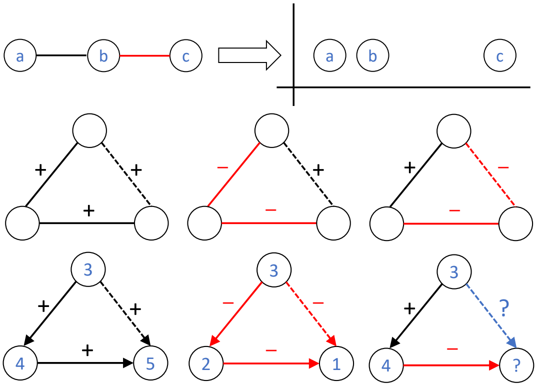
Social balance theory[24]**:**

- SiNE[25]
- SIGNet[26]
- SIDE[27]

Social status theory[28]**:**

- BESIDE[29]

Network transformation[30]



[24] Heider. Attitudes and cognitive organization. J. Psychol.'46.
[25] Wang et al. Signed network embedding in social media. SDM'17.
[26] Islam et al. Signet: Scalable embeddings for signed networks. PAKDD'18.
[27] Kim et al. Side: representation learning in signed directed networks. WebConf'18.
[28] Guha et al. Propagation of trust and distrust. WebConf'04.
[29] Chen et al. "Bridge" enhanced signed directed network embedding. CIKM'18.
[30] Javari et al. ROSE: Role-based signed network embedding. WebConf'20.

Existing methods:

- Only consider signed similarity order: positive $>$ negative
- Leave topological similarity order (connected vs disconnected) to unsigned embedding methods

[31] Leskovec et al. Predicting positive and negative links in online social networks. WebConf'10.
[32] Chiang et al. Exploiting longer cycles for link prediction in signed networks. CIKM'11.
[22] Beigi et al. Signed link prediction with sparse data: The role of personality information. WebConf'19.

Existing methods:

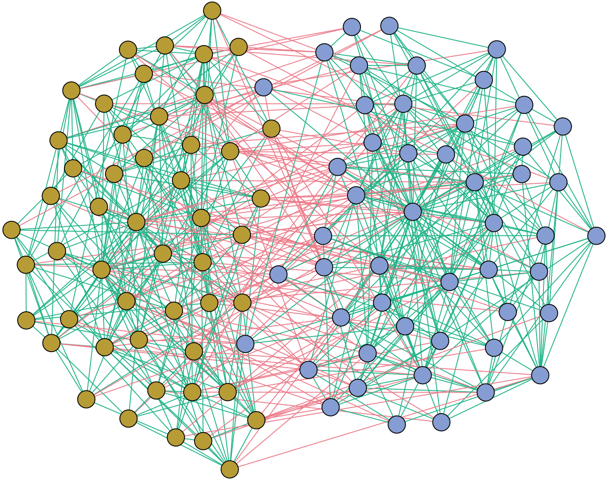
- Only consider signed similarity order: positive $>$ negative
- Leave topological similarity order (connected vs disconnected) to unsigned embedding methods

- Works for edge sign prediction[31,32] (existence of edges known)
- In signed link prediction[22] (existence of edges unknown): two-step
  1. Apply unsigned methods to predict the existence of edges
  2. Predict the signs on those pre-predicted edges

[31] Leskovec et al. Predicting positive and negative links in online social networks. WebConf'10.
[32] Chiang et al. Exploiting longer cycles for link prediction in signed networks. CIKM'11.
[22] Beigi et al. Signed link prediction with sparse data: The role of personality information. WebConf'19.

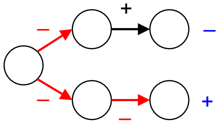
Signed link prediction in polarized graphs[2]

- Intra-community edges: dense, positive
- Inter-community edges: sparse, negative (almost impossible to predict for unsigned methods due to sparsity)
- Are real-world signed graphs often polarized?



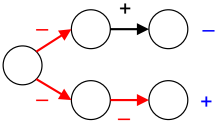
[2] **Huang**, Silva, Singh. Signed embedding for polarized graphs. Working draft.

A polarization measure



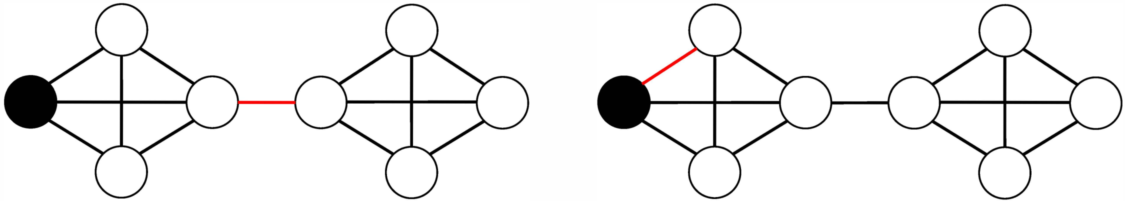
- Unsigned random-walk transition: $|M|_{ij} = \sum_{\text{all paths } l \text{ between } i \text{ and } j} \text{Prob}(l)$
- Signed random-walk transition: $M_{ij} = \sum_{\text{all paths } l \text{ between } i \text{ and } j} \text{Prob}(l)\text{Sign}(l) \leftarrow$ from balance theory

A polarization measure



- Unsigned random-walk transition: $|M|_{ij} = \sum_{\text{all paths } l \text{ between } i \text{ and } j} \text{Prob}(l)$
- Signed random-walk transition: $M_{ij} = \sum_{\text{all paths } l \text{ between } i \text{ and } j} \text{Prob}(l)\text{Sign}(l)$ $\leftarrow$ from balance theory

- Key observation: $|M|_{:i}$ and $M_{:i}$ are highly correlated if i is polarized.
- Define $\text{Pol}(i) = \text{corr}(|M|_{:i}, M_{:i})$, and $\text{Pol}(G) =$ average of $\text{Pol}(i)$.

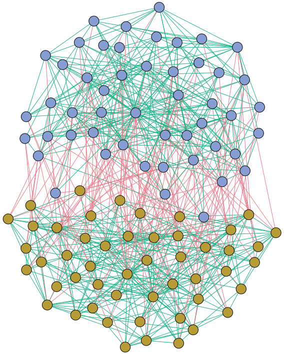


Least polarized politicians in U.S. Congress

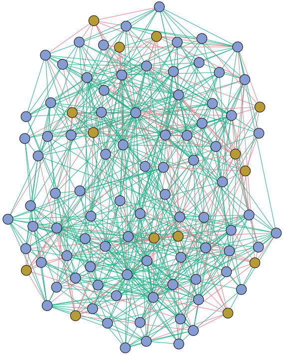
- Henry Cuellar (Pol $= -0.65$), Democrat, “moderate-centrist”, voted with President Trump 75% of the time[33]
- Jane Harman (Pol $= -0.54$), Democrat, “centralist”, “the best Republican in the Democratic Party”[34]

[33] Malone. A Q&A with the House Democrat who’s voted with trump 75 percent of the time. FiveThirtyEight’17.
[34] Skelton. California and the west: in the ring, with contenders for governor. LA Times’98.

Polarization of real-world graphs



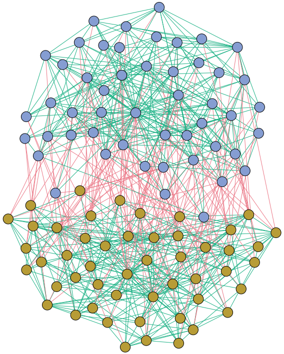
LFR[35]-polarized



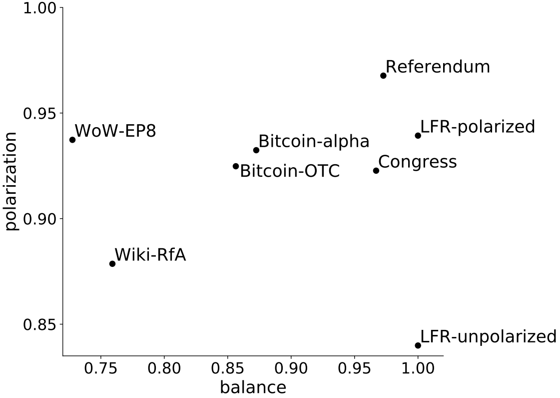
LFR-unpolarized

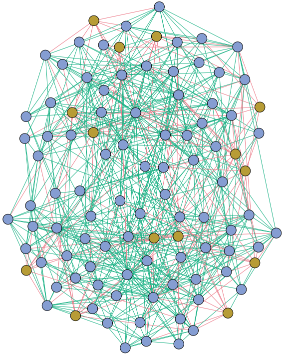
[35] Lancichinetti et al. Benchmark graphs for testing community detection algorithms. PRE'08.

Polarization of real-world graphs



LFR[35]-polarized





LFR-unpolarized

[35]Lancichinetti et al. Benchmark graphs for testing community detection algorithms. PRE'08.

Signed random-walk similarity

- Existing methods: keep signed similarity order (positive $>$ negative)
- Proposed model: extend autocovariance for signed random-walks with full similarity order (positive $>$ disconnected $>$ negative)
- Predict polarized inter-community edges with most dissimilar pairs

Signed random-walk similarity

- Existing methods: keep signed similarity order (positive > negative)
- Proposed model: extend autocovariance for signed random-walks with full similarity order (positive > disconnected > negative)
- Predict polarized inter-community edges with most dissimilar pairs

Table: Average dot product similarity between embeddings of node pairs.

	Wiki-RfA	Bitcoin-OTC	Bitcoin-alpha	Congress	WoW-EP8	Referendum
Positive	13.11	17.02	18.09	21.54	1.63	4.62
Disconnected	-0.01	-0.01	-0.02	-0.28	-0.24	-0.01
Negative	-5.92	-70.95	-56.82	-27.59	-2.85	-9.47

Signed link prediction

- For our method: rank signed similarity and predict positive/negative edges as the most similar/dissimilar pairs
- For baselines: train two classifiers on the concatenated node embeddings for positive and negative edges separately
- Knowledge on the existence of edges unavailable

Signed link prediction

Table: Our method significantly outperforms baselines in signed link prediction.

Precision @ 100%		SIGNet	SIDE	BESIDE	SiNE	ROSE	Our work
Wiki-RfA	Positive	0.0028	0.0002	0.0056	0.0098	0.0379	**0.1795**
	Negative	0.0000	0.0043	0.0016	0.0007	**0.0132**	0.0112
Bitcoin-OTC	Positive	0.0008	0.0003	0.0325	0.0058	0.0494	**0.0935**
	Negative	0.0000	0.0045	0.0076	0.0030	0.0227	**0.2727**
Bitcoin-alpha	Positive	0.0000	0.0004	0.0259	0.0040	0.0626	**0.1181**
	Negative	0.0000	0.0032	0.0226	0.0097	0.0226	**0.0903**
Congress	Positive	0.0000	0.0000	0.0125	0.0000	0.0250	**0.1250**
	Negative	0.0000	0.0000	0.0000	0.0400	0.0000	**0.1600**
WoW-EP8	Positive	0.2104	0.0949	0.1348	0.1831	0.4060	**0.6201**
	Negative	0.0361	0.0749	0.0553	0.0395	0.1221	**0.1475**
Referendum	Positive	0.0002	0.0465	0.0413	0.0440	0.0854	**0.1900**
	Negative	0.0000	0.0020	0.0056	0.0000	0.0234	**0.0429**

Signed link prediction with unsigned link prediction knowledge

- Combine the ranking scores from each method with unsigned reconstructed autocovariance similarity
- Train two classifiers to predict positive/negative edges with combined scores
- Knowledge on the existence of edges available

Signed link prediction with unsigned link prediction knowledge

Table: Our method outperforms baselines, especially for (polarized) negative edges.

Precision @ 100%		SIGNet	SIDE	BESIDE	SiNE	ROSE	Our work
Wiki-RfA	Positive	0.1700	0.1694	0.1722	0.1726	0.1710	**0.1810**
	Negative	0.0290	0.0289	0.0308	0.0314	0.0335	**0.0613**
Bitcoin-OTC	Positive	0.0811	0.0824	0.0882	0.0805	0.0844	**0.0954**
	Negative	0.1939	0.2030	0.2242	0.2152	0.1894	**0.2848**
Bitcoin-alpha	Positive	0.1053	0.1113	0.0994	0.1018	0.1145	**0.1157**
	Negative	0.1194	0.1226	0.1194	0.1161	0.1065	**0.1387**
Congress	Positive	0.1000	0.1125	0.0875	0.1000	**0.1375**	0.1000
	Negative	0.0400	0.0400	0.0000	0.0000	0.0000	**0.1600**
WoW-EP8	Positive	0.4454	0.4624	0.4553	0.4605	0.6014	**0.6163**
	Negative	0.0911	0.1367	0.1288	0.1233	0.1644	**0.2539**
Referendum	Positive	0.1746	0.1775	0.1719	0.1706	0.1714	**0.1903**
	Negative	0.0461	0.0326	0.0942	0.0679	0.0457	**0.1132**

Contributions

- Design a polarization measure for nodes in/and signed graphs.
- Identify the problem of past work in polarized signed link prediction.
- Propose a signed embedding model that solves the problem and significantly outperforms SOTA in various real-world signed graphs.

Future work

- Extending other types of similarity to signed graphs (e.g., PMI).
- A model that leverages the polarization measure to guide the use of signed and unsigned similarity for signed link prediction.

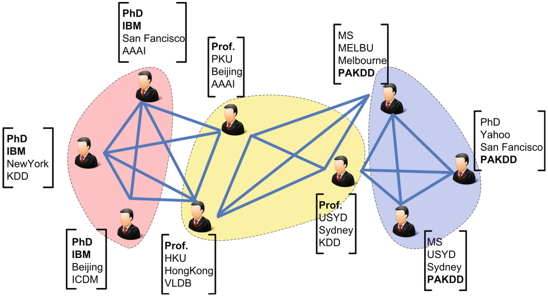
Outline

Attributed graphs

- Citation networks[36]:
  e.g., Cora, Pubmed
- Wikipedia networks[37]:
  e.g., Chameleon, Squirrel

Embedding objectives

- Unsupervised[38]:
  e.g., topological similarity
- (Semi-)supervised[39]:
  e.g., node classification



[36] Sen et al. Collective classification in network data. AI Magazine'08.
[37] Rozemberczki et al. Multi-scale attributed node embedding. J. Complex Netw.'21.
[38] Kipf, Welling. Variational graph auto-encoders. arXiv'16.
[39] Kipf, Welling. Semi-supervised classification with graph convolutional networks. arXiv'16.

Graph Neural Networks (GNN)

- Rationale: learn to aggregate attributes from neighbors.

Aggregation

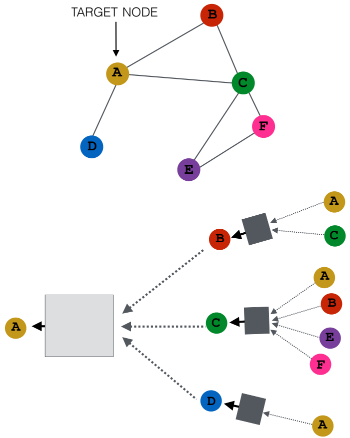
- GCN[39], GraphSAGE[40]

Edge weighting

- AGNN[41], GAT[42], GraphSAGE

Multi-hop neighborhood

- LanczosNet[43], PPNP[44]



[39] Kipf, Welling. Semi-supervised classification with graph convolutional networks. arXiv'16.
[40] Hamilton et al. Inductive representation learning on large graphs. NeurIPS'17.
[41] Thekumparampil et al. Attention-based graph neural network for semi-supervised learning. arXiv'18.
[42] Veličković et al. Graph attention networks. ICLR'18.
[43] Liao et al. Lanczosnet: Multi-scale deep graph convolutional networks. ICLR'19.
[44] Klicpera et al. Predict then propagate: Graph neural networks meet personalized pagerank. ICLR'19.

Challenges:

- Stacking too few layers can't capture the multi-hop information.
- Stacking too many layers leads to oversmoothing[45,46].

Solution: learn to aggregate multi-hop neighborhood[3]

[45] Li et al. Deeper insights into graph convolutional networks for semi-supervised learning. AAAI'18
[46] Xu et al. Representation learning on graphs with jumping knowledge networks. ICML'18.
[3] Ye, **Huang**, Singh. Krylov graph convolutional networks. Under review.

Simple Graph Convolution[47]:

$$H^{(k)} = \sigma(\widetilde{M}^k X \Theta)$$

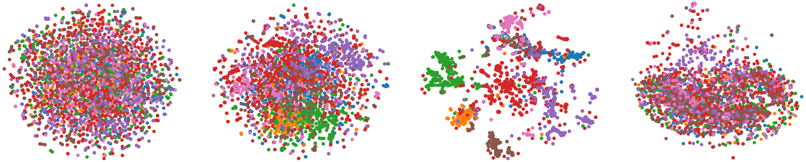
$H^{(k)}$: kth order embedding, X: node attributes, Θ: trainable weights
$\widetilde{M}$: random-walk transition with added self-loops

[47] Wu et al. Simplifying graph convolutional networks. ICML'19.

Simple Graph Convolution[47]:

$$H^{(k)} = \sigma(\widetilde{M}^k X \Theta)$$

$H^{(k)}$: kth order embedding, X: node attributes, Θ: trainable weights
$\widetilde{M}$: random-walk transition with added self-loops



(a) $k = 0$ (b) $k = 1$ (c) $k = 19$ (d) $k = 10000$

Figure: Visualization of embedding without training for Cora. Colors denote actual classes. A good choice of neighborhood helps reveal the cluster structure.

[47]Wu et al. Simplifying graph convolutional networks. ICML'19.

Weighting neighborhoods

- Need to assign right weights to different orders of neighborhoods.
- Crucial for semi-supervised learning in sparsely labeled graphs.
- Weights should be adaptive to different datasets.

Weighting neighborhoods

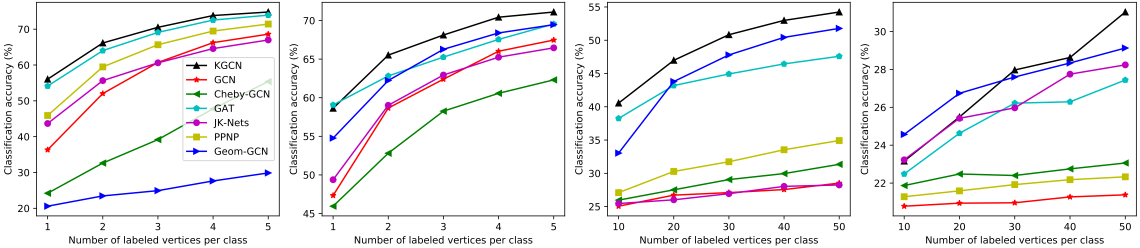
- Need to assign right weights to different orders of neighborhoods.
- Crucial for semi-supervised learning in sparsely labeled graphs.
- Weights should be adaptive to different datasets.

Krylov Graph Convolutional Network

$$H = \sigma(\sum_{k=0}^{K} \alpha_k \widetilde{M}^k X\Theta)$$

- Weights for different neighborhoods are learned for each dataset.
- Krylov tensor $(X\Theta, \widetilde{M}X\Theta, \dots, \widetilde{M}^K X\Theta)$ can be computed iteratively for efficiency, since $\widetilde{M}$ is sparse.
- This procedure can also be viewed as the power iteration for $\widetilde{M}$.

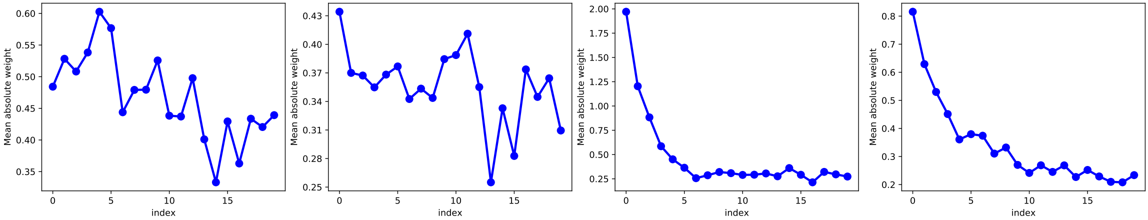
Semi-supervised node classification for sparsely labeled graphs



(a) Cora (b) Pubmed (c) Chameleon (d) Squirrel

Figure: Our method (KGCN) outperforms the baselines in almost all cases.

Learned neighborhood weights



(a) Cora (b) Pubmed (c) Chameleon (d) Squirrel

Figure: Mean neighborhood weights (over 30 splits) are different across datasets. Note that Cora and Pubmed have larger weights for high-order neighborhoods compared to Chameleon and Squirrel.

Contributions:

- Interpret the GCN propagation rule as power iteration, with its Krylov tensor encoding multi-hop neighborhood information.
- Propose to learn the weights for combining different orders of neighborhood information, adaptive to different datasets.
- Outperforms SOTA on various sparsely labeled graphs.

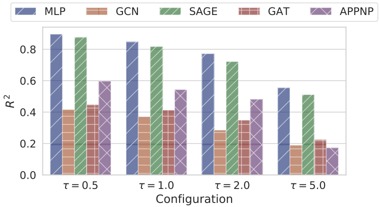
Leveraging topology and attribute information

- Internship: attributed multiplex heterogeneous graph embedding[48]
  - GCNs, GATs, random-walks, skip-grams, etc.
  - Authors provide a over-simplified implementation.

[48] Cen et al. Representation learning for attributed multiplex heterogeneous network. KDD'19.
[49] Ma et al. CopulaGNN: Towards integrating representational and correlational roles of graphs in GNNs. ICLR'21.
[50] Srinivasan, Ribeiro. On the equivalence between node embeddings and structural graph representations. ICLR'20.

Leveraging topology and attribute information

- Internship: attributed multiplex heterogeneous graph embedding[48]
  - GCNs, GATs, random-walks, skip-grams, etc.
  - Authors provide a over-simplified implementation.

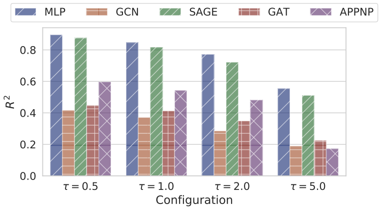


All GNNs underperform the basic MLP in regression[49].

[48] Cen et al. Representation learning for attributed multiplex heterogeneous network. KDD'19.
[49] Ma et al. CopulaGNN: Towards integrating representational and correlational roles of graphs in GNNs. ICLR'21.
[50] Srinivasan, Ribeiro. On the equivalence between node embeddings and structural graph representations. ICLR'20.

Leveraging topology and attribute information

- Internship: attributed multiplex heterogeneous graph embedding[48]
  - GCNs, GATs, random-walks, skip-grams, etc.
  - Authors provide a over-simplified implementation.



All GNNs underperform the basic MLP in regression[49].

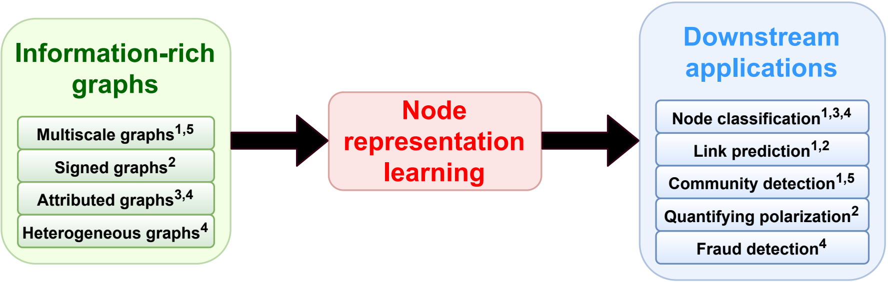
- Need better ways to use topological information in GNNs, not just as the computational graph for aggregation.
- As an attempt, CopulaGNN[49] adds correlational roles for node-level regression.
- Topological information is even more important for link prediction[50].

[48] Cen et al. Representation learning for attributed multiplex heterogeneous network. KDD'19.
[49] Ma et al. CopulaGNN: Towards integrating representational and correlational roles of graphs in GNNs. ICLR'21.
[50] Srinivasan, Ribeiro. On the equivalence between node embeddings and structural graph representations. ICLR'20.

Outline

Conclusions: extracting key information for applications

- Heterogeneous degree distribution for link prediction[1]
- Full similarity order for polarized signed link prediction[2]
- Adaptive neighborhood weighting for sparsely labeled node classification[3]



[1] **Huang**, Silva, Singh. A broader picture of random-walk based graph embedding. KDD'21.
[2] **Huang**, Silva, Singh. Signed embedding for polarized graphs. Working draft.
[3] Ye, **Huang**, Singh. Krylov graph convolutional networks. Under review.
[4] **Huang**. Graph-based fraud detection in Kindle Direct Publishing. Amazon internship report.
[5] **Huang***, Kondapaneni*, Silva, Singh. Multiscale community detection based on PMI. Ongoing.

Plan

Jun/21: Start working on leveraging topology and attributes

Aug/21: Finish paper on signed embedding

Oct/21: Finish paper on PMI clustering

Jun/22: Proposal

...

?/23: Defense