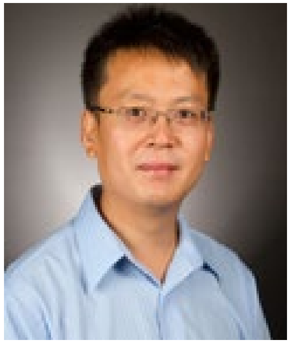
ImVerde: Vertex-Diminished Random Walk for Learning Imbalanced Network Representation

Presenter: Jun Wu

joint work with

Dr. Jingrui He and **Dr. Yongming Liu**



Arizona State University
Jingrui.he@asu.edu

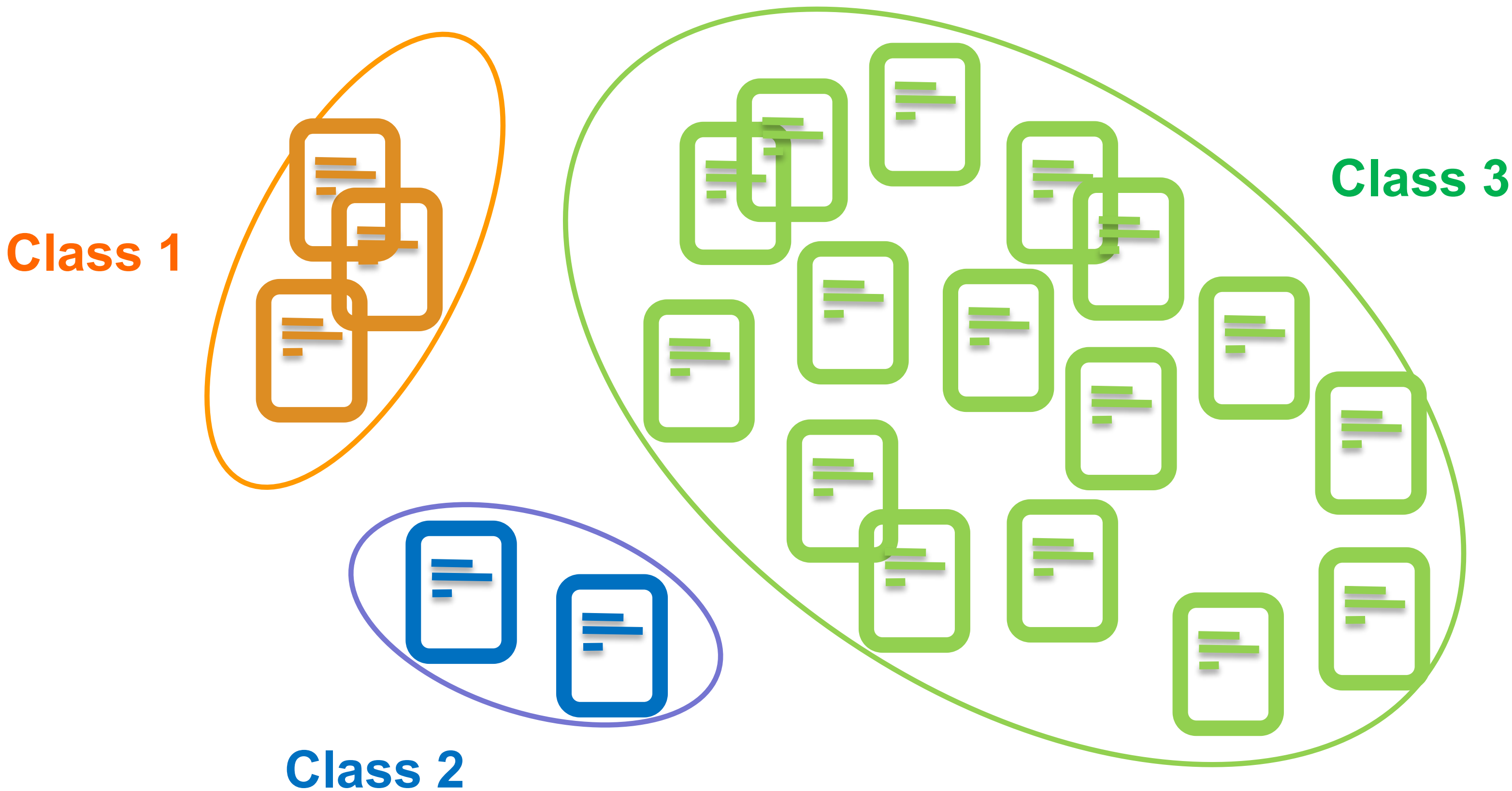
Arizona State University
Yongming.Liu@asu.edu

Roadmap

- Motivation
- Proposed *VDRW* algorithm
- Proposed *ImVerde* framework
- Experiments
- Conclusion

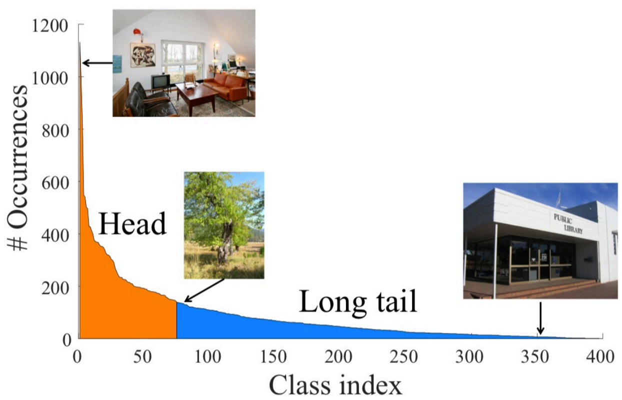
Imbalanced data is everywhere…

- The data distribution from different classes are significantly skewed.

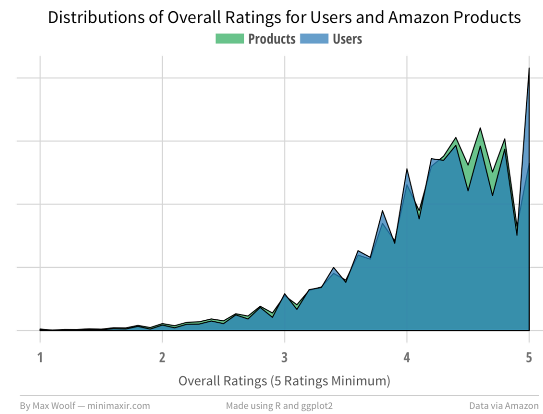


Imbalanced data is everywhere…

Grid-like structure

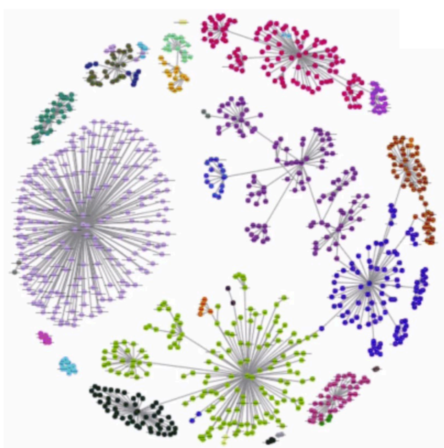


SUN-397 image data set



Amazon reviews

Graph structure



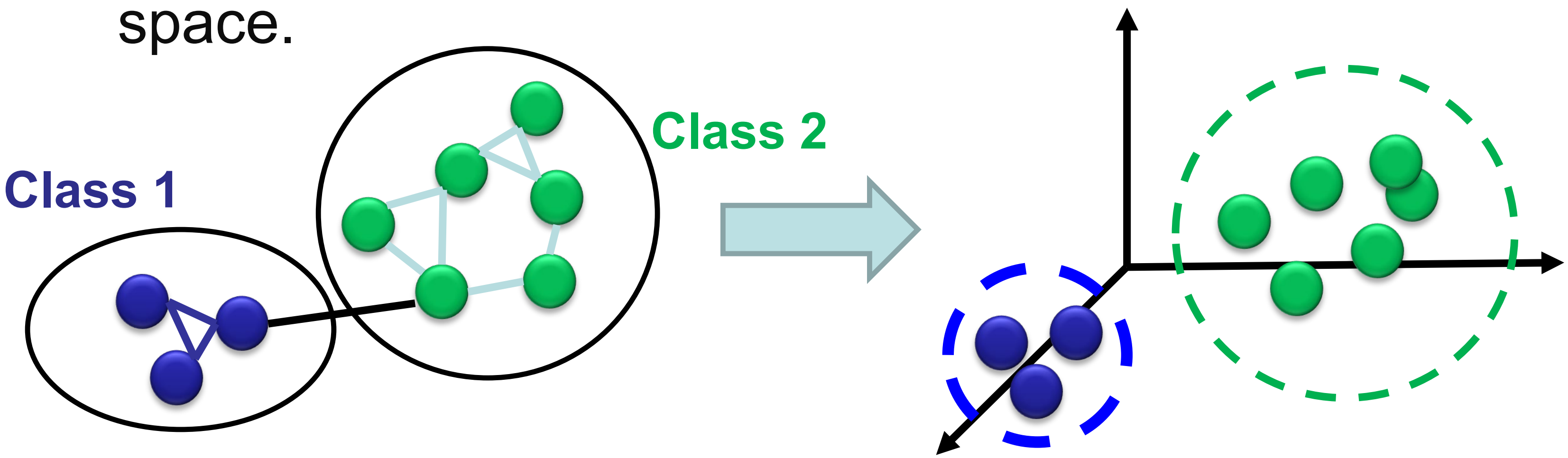
Social network

Online transaction

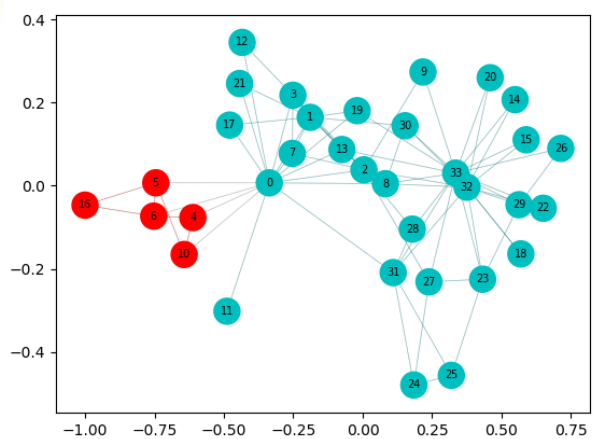
Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. "Learning to model the tail." In *Advances in Neural Information Processing Systems*, pp. 7029-7039. 2017.

Representation learning

- Problem definition
  - Input: (i) a (directed or undirected) network $G = (V, E)$, (ii) imbalanced class labels for nodes in V
  - Output: a low-dimensional vector representation for each node $v \in V$, so that the minority class is separated from majority class in the embedding feature space.



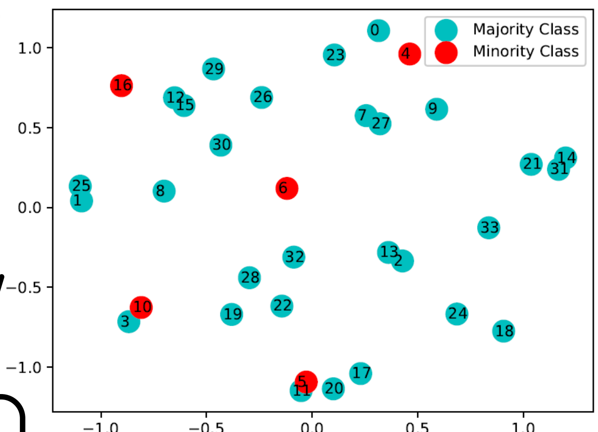
Imbalanced network embedding



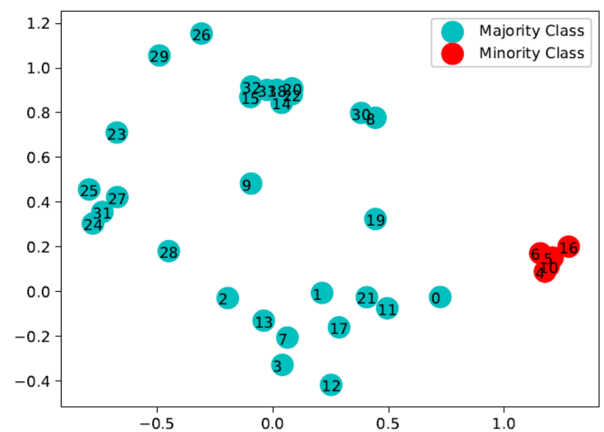
Karate network

Network embedding

Network embedding



Planetoid



ImVerde

Visualization via t-SNE

Visualization via t-SNE

Wayne W. Zachary. "An information flow model for conflict and fission in small groups." *Journal of anthropological research*33, no. 4 (1977): 452-473.

Roadmap

- Motivation
- Proposed *VDRW* algorithm
- Proposed *ImVerde* framework
- Experiments
- Conclusion

Vertex-Diminished Random Walk

- Intuition: the probability of a transition to one node would decrease each time when it is visited.

$$P(Y_{t+1} = j|\mathcal{F}_t) = \frac{R_{Y_t,j} f(S_j(t))}{\sum_i R_{Y_t,i} f(S_i(t))} \tag{1}$$

where

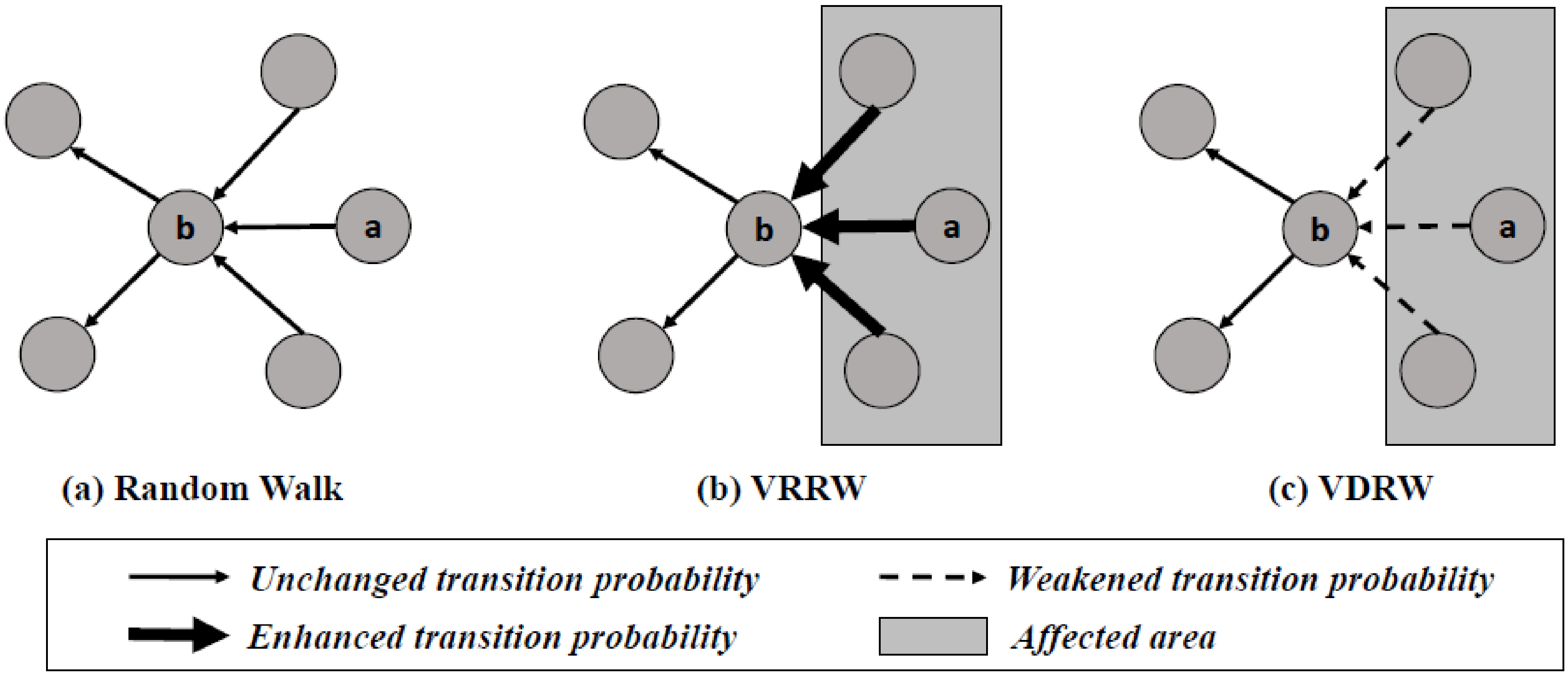
$$f\big(S_j(t)\big) = \alpha^{S_j(t)} \tag{2}$$

R_{ij}: transition probability from node i to node j
$S_j(t)$: the number of visits to node j up to time t
$f(\cdot)$: a visiting function based on $S_j(t)$
α: hyper-parameter to be specified

Robin Pemantle. "Vertex-reinforced random walk." *Probability Theory and Related Fields* 92, no. 1 (1992): 117-136.

Vertex-Diminished Random Walk

- Intuition: the probability of a transition to one node would decrease each time when it is visited.



$\boxed{f\left(S_j(t)\right) = 1}$ $\boxed{f\left(S_j(t)\right) = S_j(t)}$ $\boxed{f\left(S_j(t)\right) = \alpha^{S_j(t)}}$

Robin Pemantle. "Vertex-reinforced random walk." *Probability Theory and Related Fields* 92, no. 1 (1992): 117-136.

Vertex-Diminished Random Walk

- Convergence analysis

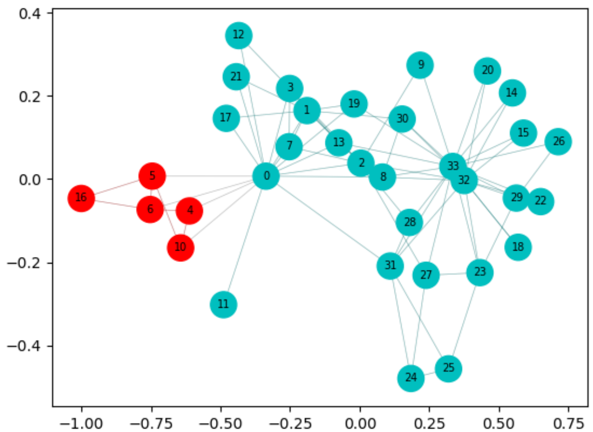
$$V_i(t) = \frac{f(S_i(t))}{\sum_i f(S_i(t))}$$

Theorem 1: With probability one, $dist(V(t), \mathcal{C} \cup \mathcal{C}_0) \to 0$ where

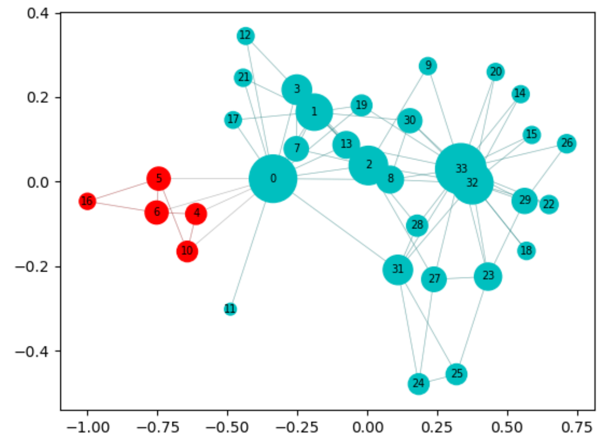
$$dist(x, A) = \inf\{|x - y| : y \in A\}$$

Corollary 3: If all the off-diagonal entries of R are positive and all the principal minors of R are invertible, $V(t)$ converges almost surely.

Vertex-Diminished Random Walk



Karate Network

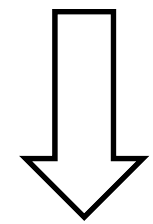


Random Walk

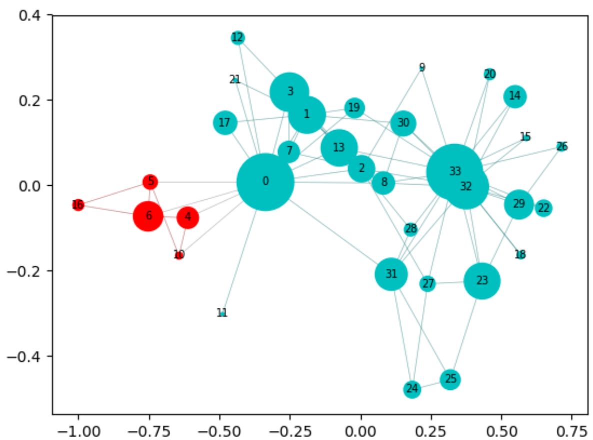
Minority class

Majority class

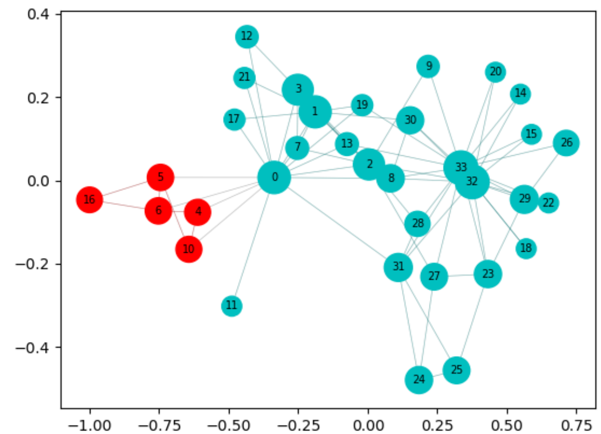
Larger circle



Higher visited times



Vertex-Reinforced Random Walk



Vertex-Diminished Random Walk

Wayne W. Zachary. "An information flow model for conflict and fission in small groups." *Journal of anthropological research*33, no. 4 (1977): 452-473.

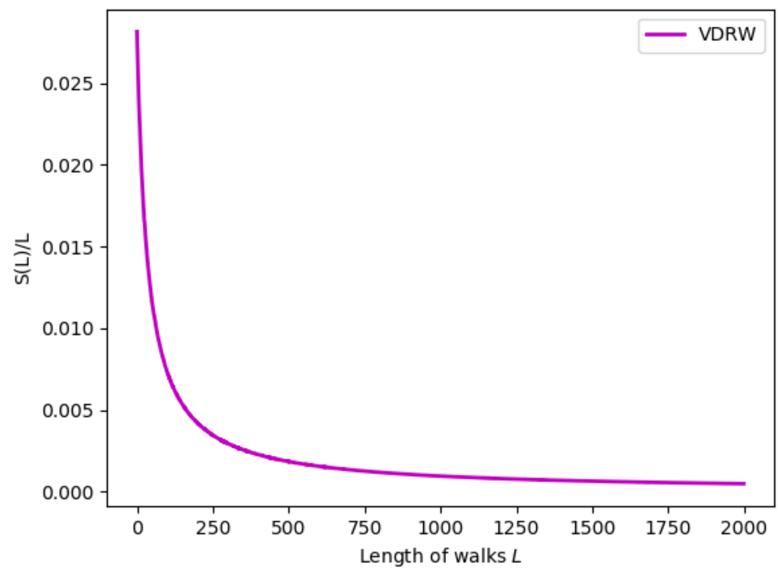
Vertex-Diminished Random Walk

Walking 10 steps for 100 times

Walking 2000 steps once

Minority class　Majority class

Accuracy of node-context pairs



Convergence analysis

Case study on the Karate network

Wayne W. Zachary. "An information flow model for conflict and fission in small groups." *Journal of anthropological research*33, no. 4 (1977): 452-473.

Roadmap

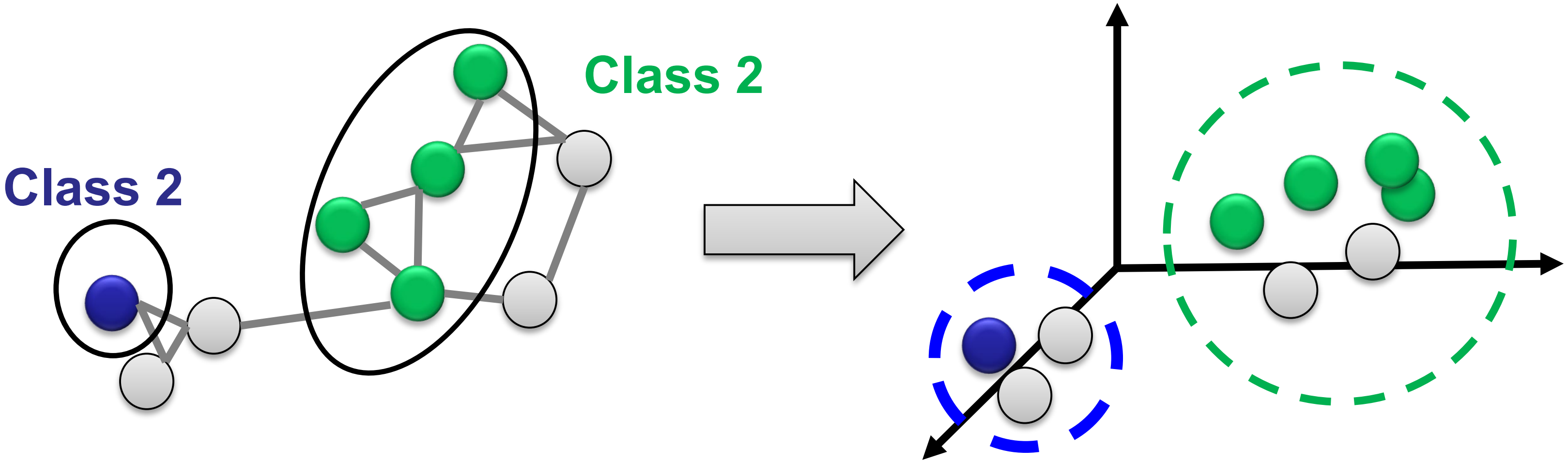
- Motivation
- Proposed *VDRW* algorithm
- Proposed *ImVerde* framework
- Experiments
- Conclusion

Imbalanced network embedding

- The loss of our semi-supervised framework:

$$\mathcal{L} = \mathcal{L}_s + \lambda \mathcal{L}_u$$

  - $\mathcal{L}_s$: supervised loss of predicting the label
  - $\mathcal{L}_u$: unsupervised loss of predicting the graph context
    - Assumption: nodes with the same context would have the similar embeddings



Imbalanced network embedding

- Unsupervised loss

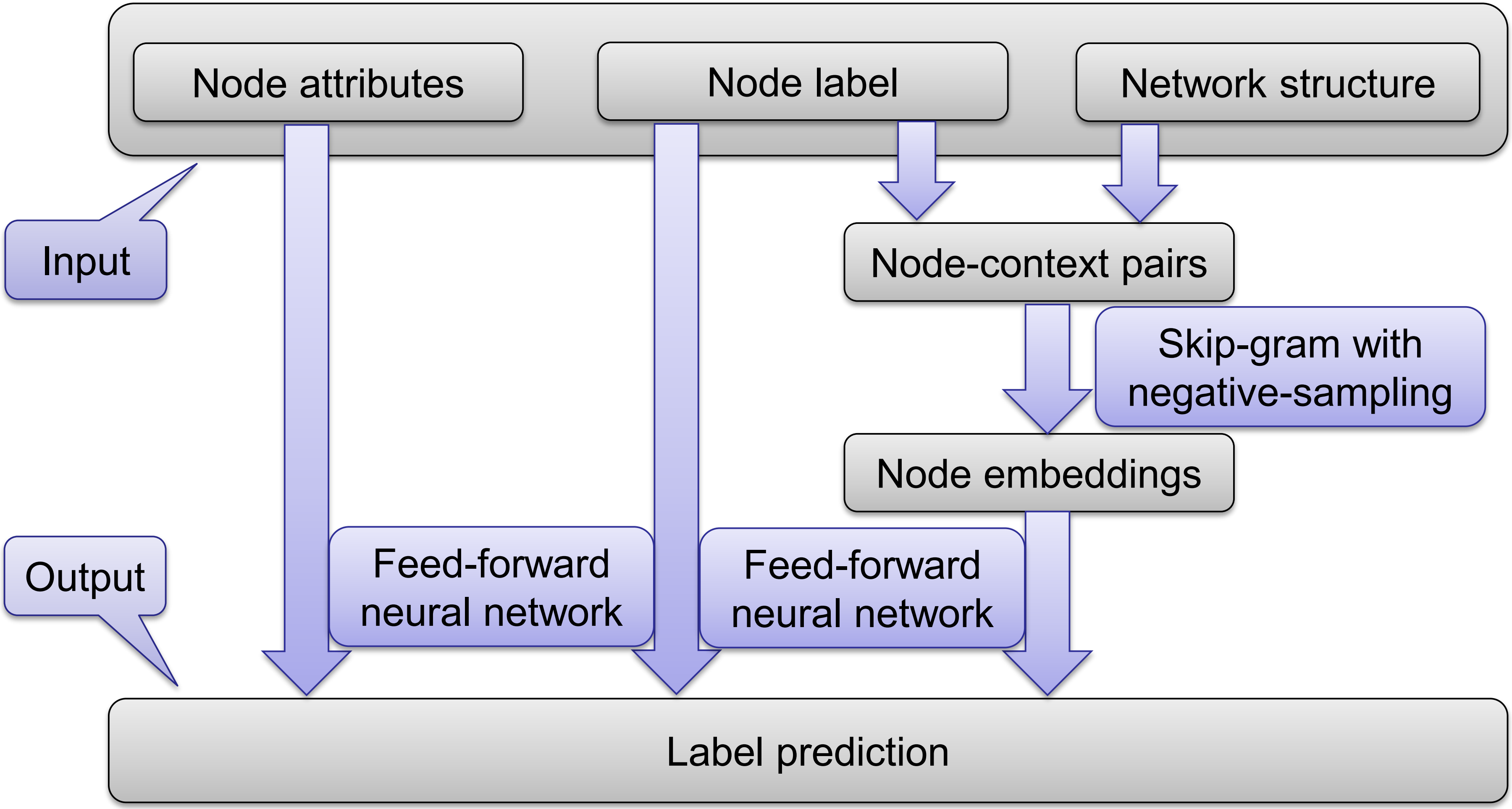
$$\mathcal{L}_u = -\sum_{(v_i, v_c)} \left[\log \sigma(w_c^T e_i) + k \cdot \mathbb{E}_{v_n \sim P_n(v_c)} \log \sigma(-w_n^T e_i)\right]$$

> **(v_i, v_c): a node-context pair of node i and its corresponding context c**
> **e_i: embedding of node i**
> **w_c: representation of context c**

- Supervised loss (cross-entropy error)

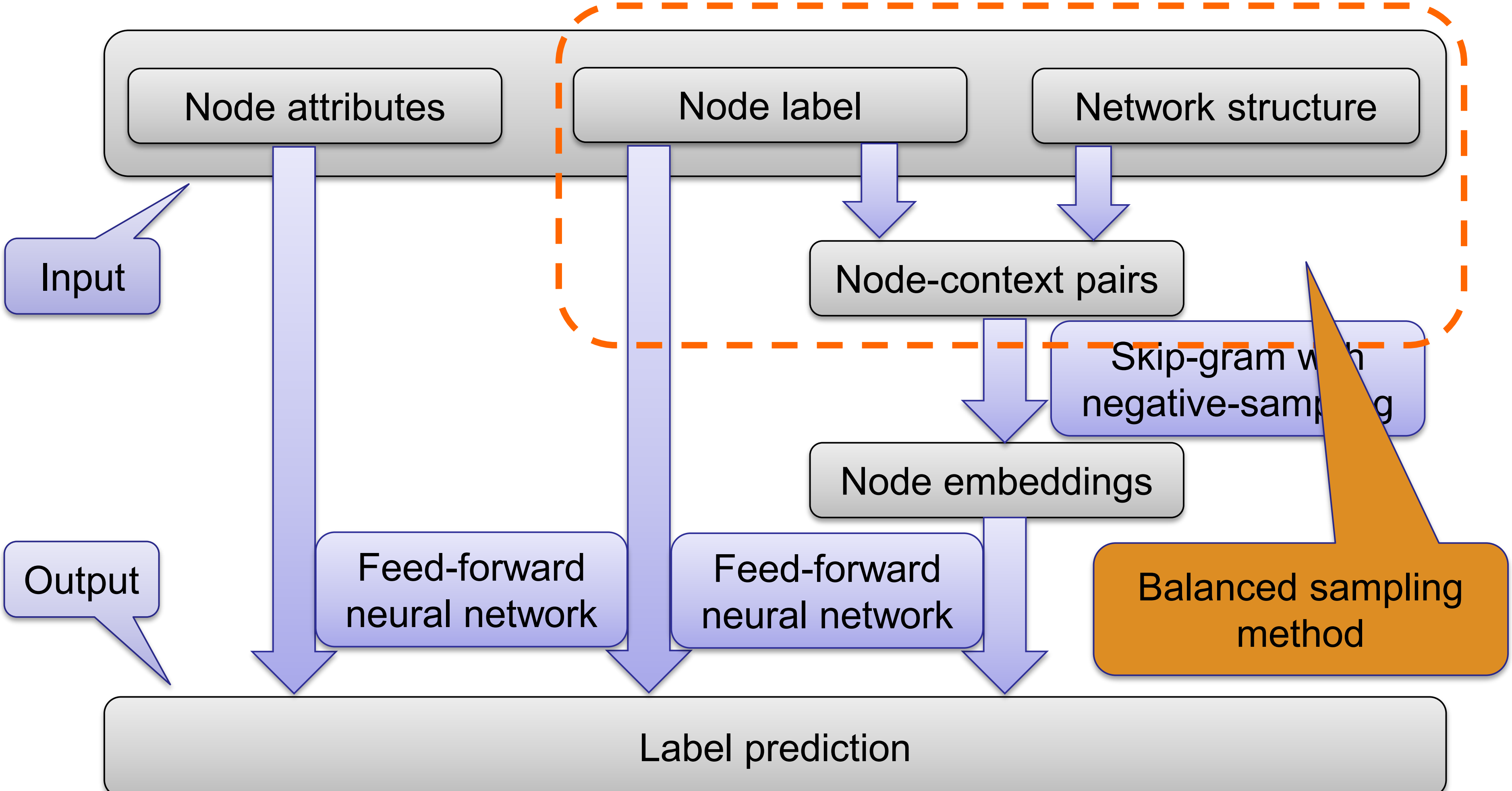
$$\mathcal{L}_s = -\frac{1}{l}\sum_{i=1}^{l} \log \frac{\exp\left([h^{l1}(x_i)\; h^{l2}(e_i)]^T y_i\right)}{\sum_{y'} \exp([h^{l1}(x_i)\; h^{l2}(e_i)]^T y')}$$

Imbalanced network embedding



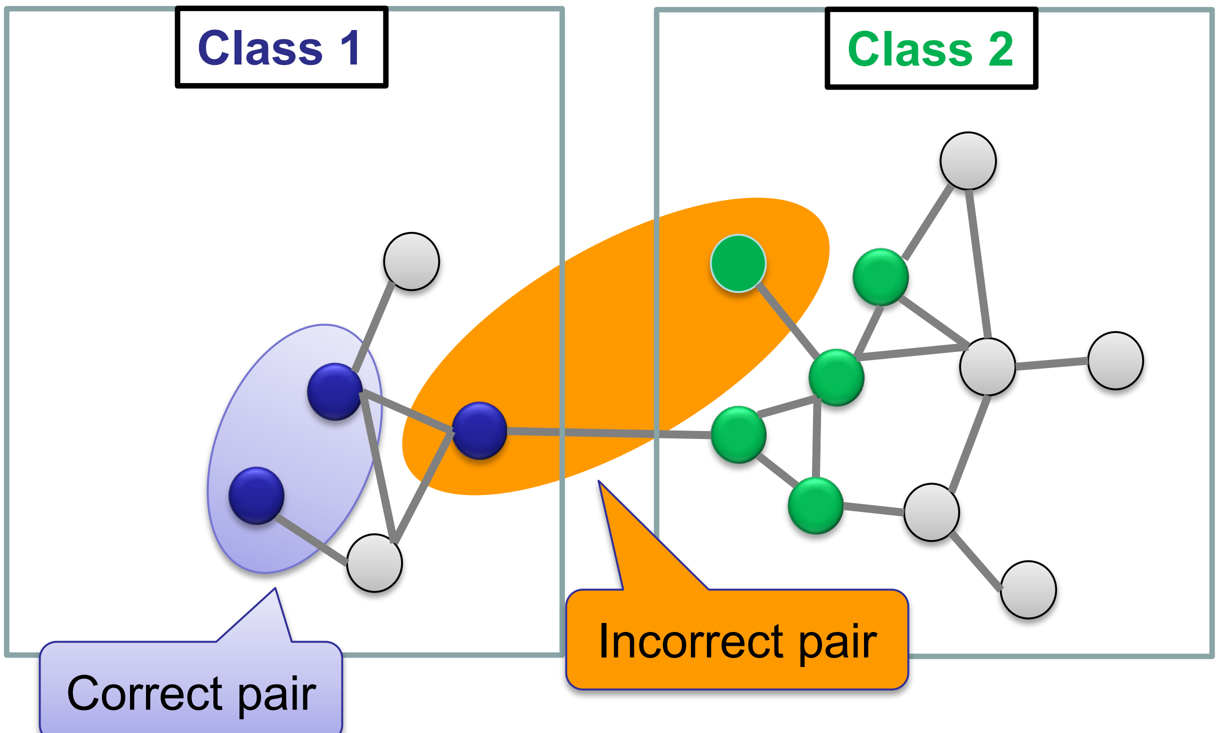
Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. "Revisiting semi-supervised learning with graph embeddings." *In ICML* (2016).

Imbalanced network embedding



Imbalanced network embedding

- Key intuition: node-context pairs within the same class are crucial for learning node representation in the minority class.



Context sampling

- Node-context pair sampling:
  - Label information: nodes with the same label would be seen as the contexts for each other
  - Network structure: short VDRW can explore the context for each node within the class

Context sampling

- Node-context pair sampling:
  - Label i… would be see…
  - Netwo… he contex…

Algorithm 2 *Context Sampling*

Input: Graph G with transition matrix $\mathbf{R}$, labels $y_{1:l}$
initial node v_s
jumping probability r
length of walking sequences T

Output: Node-context pairs (v_i, v_c)

```
Initialize: Add v_s to walking path p;
for t = 0 to T − 1 do
    if Node is labeled then
        With r, select one node v with the same label;
        With 1 − r, select its neighbor v based on R;
    else
        Select its neighbor v based on R;
    end if
    Update R using VDRW;
    Update the walking path p;
end for
Return the node-context pairs sampled from path p.
```

Balanced-batch sampling

- It adopted the under-sampling method to select the seeds for mini-batch training
  - All the labeled minority instances are selected

$$|S_{min}|$$

  - Randomly select the labeled majority instances

$$|S'_{maj}| = |S_{min}|$$

  - Randomly select the unlabeled instances

$$|S_{other}| = Batchsize - |S'_{maj}| - |S_{min}|$$

Roadmap

- Motivation
- Proposed *VDRW* algorithm
- Proposed *ImVerde* framework
- Experiments
- Conclusion

Experiments

- Data sets:

	# nodes	# edges	# classes	# features
Cora	2,708	5,439	7	1,433
Citeseer	3,327	4,732	6	3,703
Pubmed	19,717	44,338	3	500
NTSB	1,473	11,784	4	4,424

- Baseline methods:
  - Planetoid
  - Node2vec
  - DeepWalk
  - SMOTE

NTSB: each node represents one accident report from National Transportation Safety Board database, and each edge reflects the similarity between two reports using TF-IDF features. Node labels correspond to the accident causes.

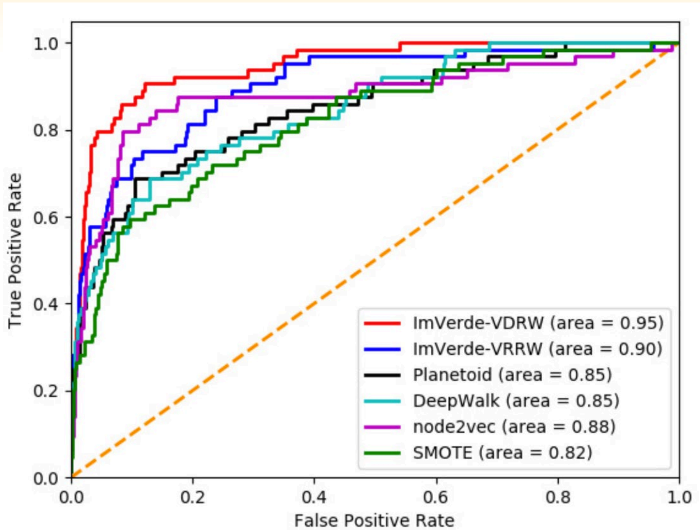
https://www.ntsb.gov/Pages/default.aspx

Binary node classification

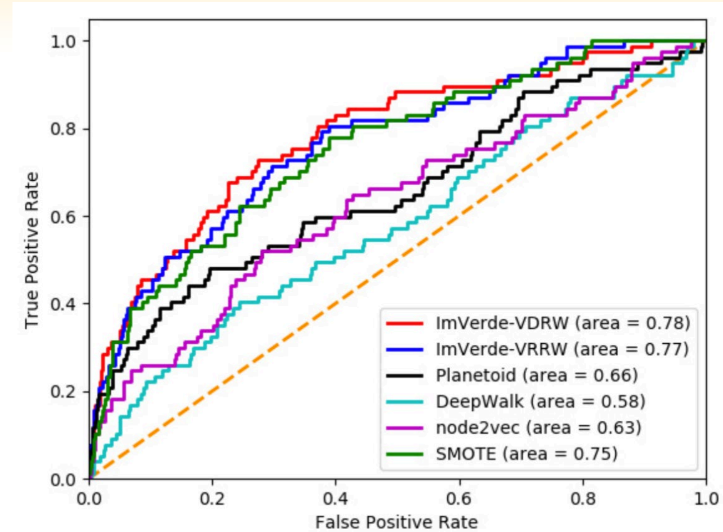
	Planetoid	DeepWalk	node2vec	SMOTE	*ImVerde-VDRW*
Cora_1	0.639	0.650	0.612	0.473	**0.720**
Cora_2	0.749	0.778	0.761	0.664	**0.852**
Cora_3	0.900	0.863	0.840	0.707	**0.910**
Cora_4	0.783	0.657	0.761	0.723	**0.826**
Cora_5	0.637	0.610	0.471	0.641	**0.769**
Cora_6	0.557	0.722	0.505	0.584	**0.805**
Cora_7	0.501	0.474	0.503	0.404	**0.657**
Citeseer_1	0.238	0.116	0.169	0.252	**0.304**
Citeseer_2	0.396	0.225	0.250	**0.550**	0.513
Citeseer_3	0.651	0.616	**0.693**	0.605	0.666
Citeseer_4	0.608	0.352	0.415	0.729	**0.732**
Citeseer_5	0.706	0.512	0.610	**0.790**	0.785
Citeseer_6	0.591	0.310	0.432	0.644	**0.651**
Pubmed_1	0.298	0.335	0.203	0.631	**0.631**
Pubmed_2	0.454	0.632	0.497	**0.669**	0.663
Pubmed_3	0.489	**0.736**	0.462	0.685	0.653

Average precision for the binary node classification (The best results are indicated in bold)

Binary node classification

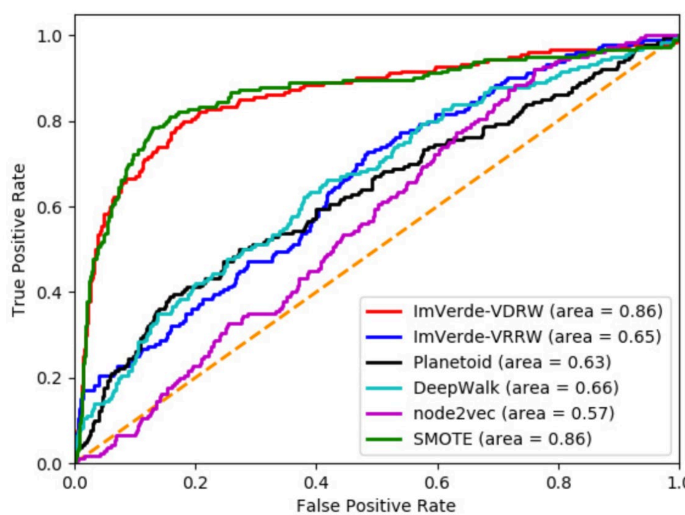


(a) Cora



(b) Citeseer

Higher is better



(c) Pubmed

The ROC curve and AUC value for node classification on (a) Cora, (b) Citeseer, (c) Pubmed data sets

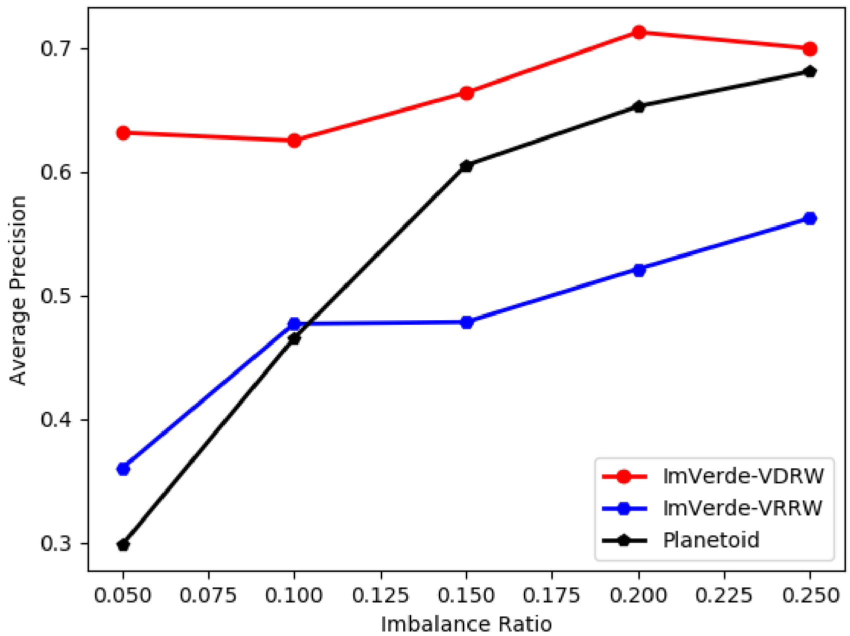
Multi-class node classification

	Accuracy
Planetoid	0.311
DeepWalk	0.458
node2vec	0.408
ImVerde-VRRW	0.416
ImVerde-VDRW	**0.460**

The accuracy of multi-class text classification on NTSB data set

Imbalance analysis

Higher is better



The average precision with different imbalance ratio on Pubmed data set

Roadmap

- Motivation
- Proposed *VDRW* algorithm
- Proposed *ImVerde* framework
- Experiments
- Conclusion

Conclusion

- Vertex-Diminished Random Walk (VDRW) is proposed, which is effective to extract the node-context pairs from imbalanced network
- A novel semi-supervised framework, *ImVerde*, is presented to learn the network representation from imbalanced data
- Experimental results on real-world data sets demonstrated the effectiveness of the proposed model

Thanks!
&
Questions?

Presenter: **Jun Wu** Email: **junwu6@asu.edu**