Machine Learning and Data Mining Models Applied to Social Analysis, Healthcare, Financial Fraud Analysis

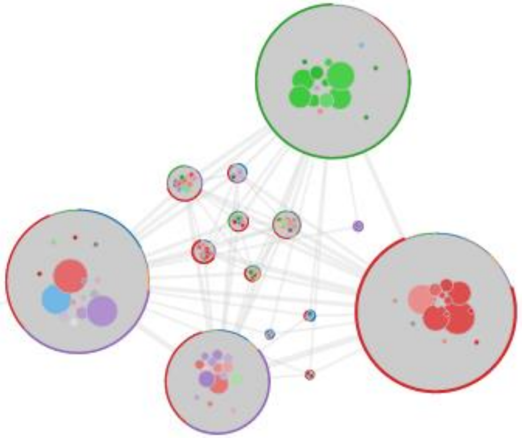
Presenter

Lecheng Zheng

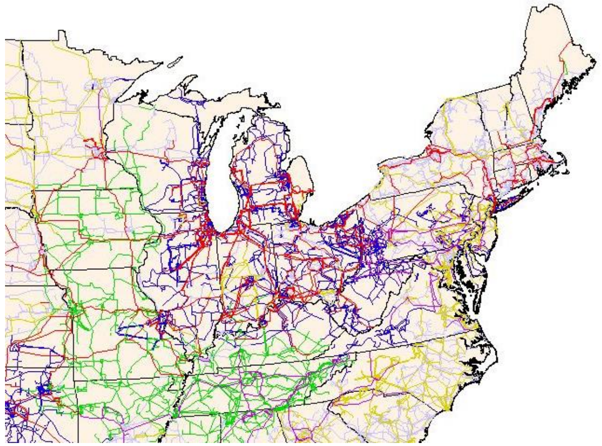
University of Illinois at Urbana-Champaign,

Email: lecheng4@illinois.edu

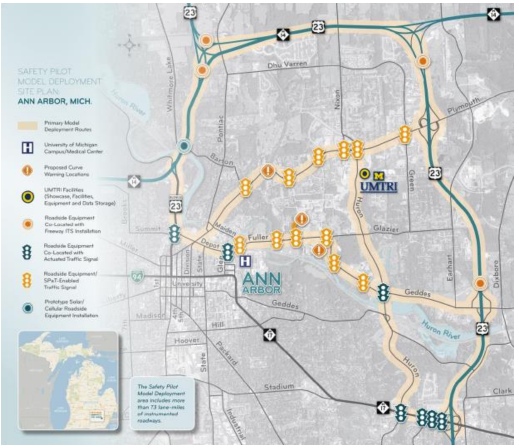
Observation: Graphs Are Everywhere!



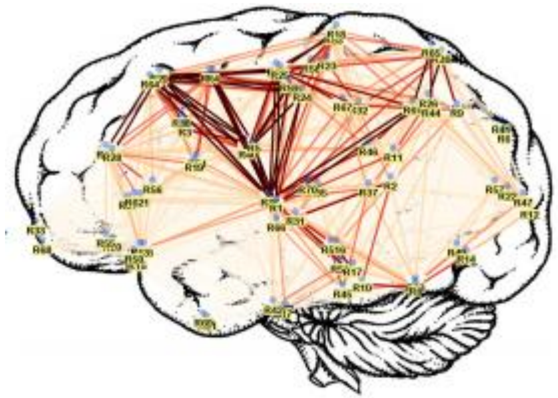
Collaboration Networks



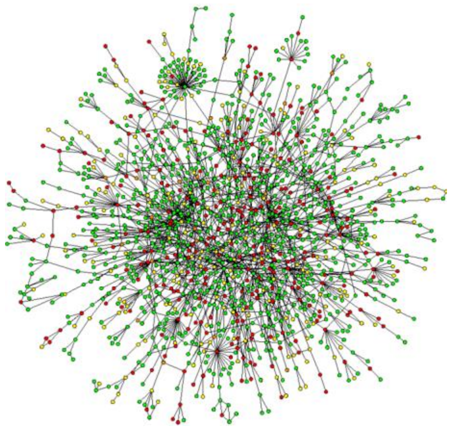
US Power Grid



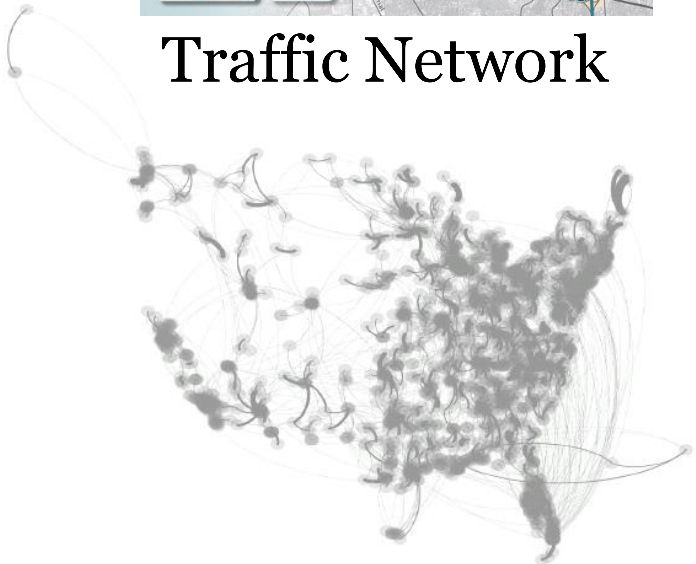
Traffic Network



Brain Networks



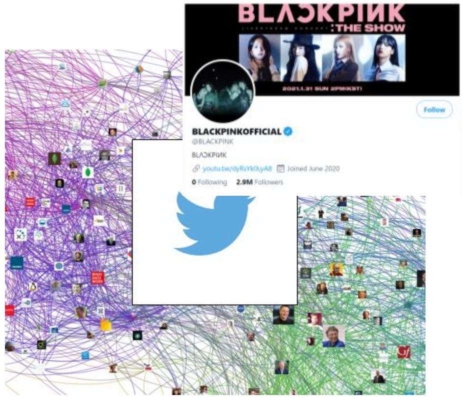
Biological Networks



Hospital Networks

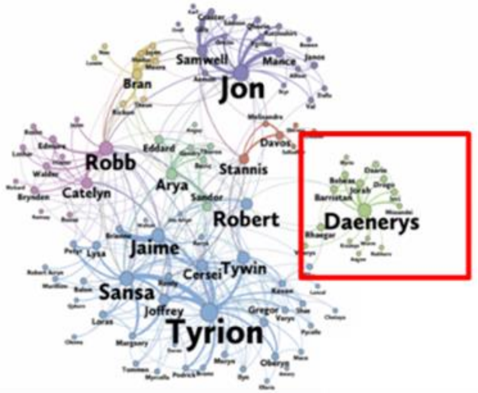
Examples of Rare Categories on Graphs

Important Rare Categories!

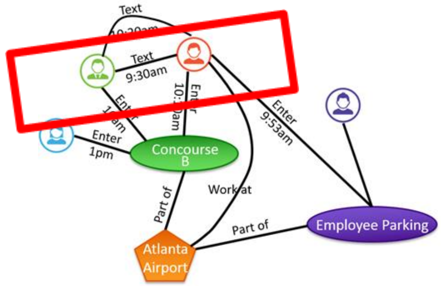


Social Community
Ex. Fans of Blackpink

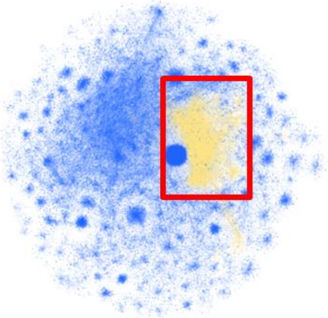
Money Laundering
Ex. Accounts involved in money laundering



Airport Insiders
Ex. gun-smuggling ring (e.g., Delta employees)



Misinformation Dissemination*
Ex: accounts disseminating disinformation in the early stage



% of the Target Examples

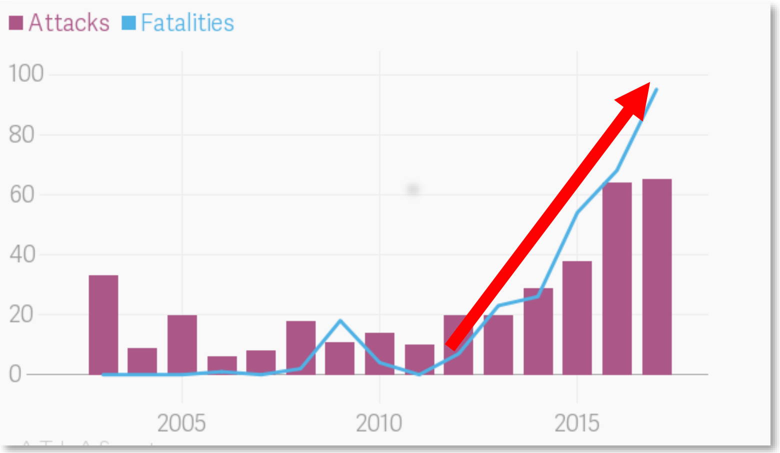
Common Category

Rare Category

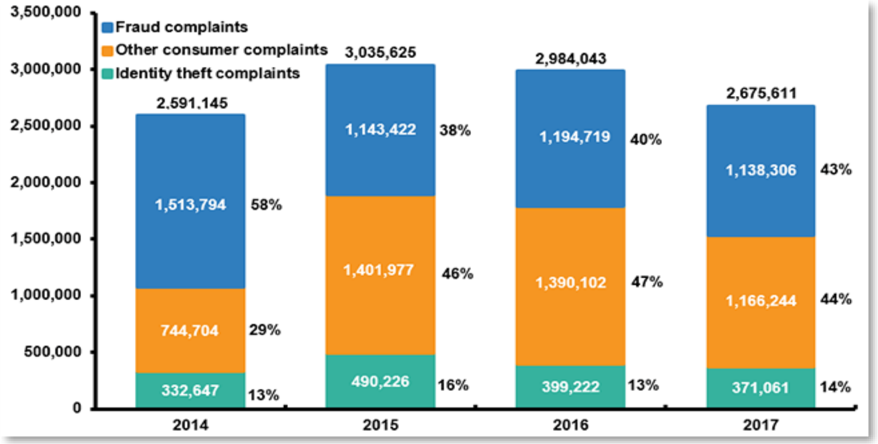
- Wang, Yu-Xiong, et al. "Learning to model the tail." NeurIPS. 2017.
- Laite, Ralph, et al. "Terrorist Group Classification of Historic Terrorist Attacks from The Global Terrorism Database." ICCSE. 2019.
- *Picture from: https://cnets.indiana.edu/blog/2016/12/21/hoaxy/

Why Do We Care (Security Domain)?

- Terrorist Events Worldwide



- Financial Fraud

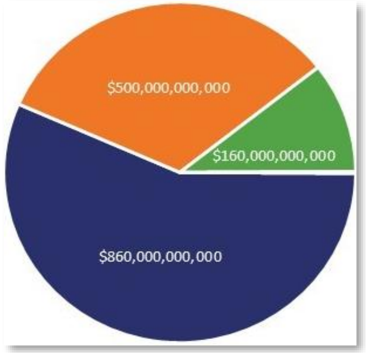


- PR Crisis

US

Allies

- Cybercrime



- Illegal online
- Markets
- Trade secret, IP theft, data trading

Cost over $3.5 billion in 2019

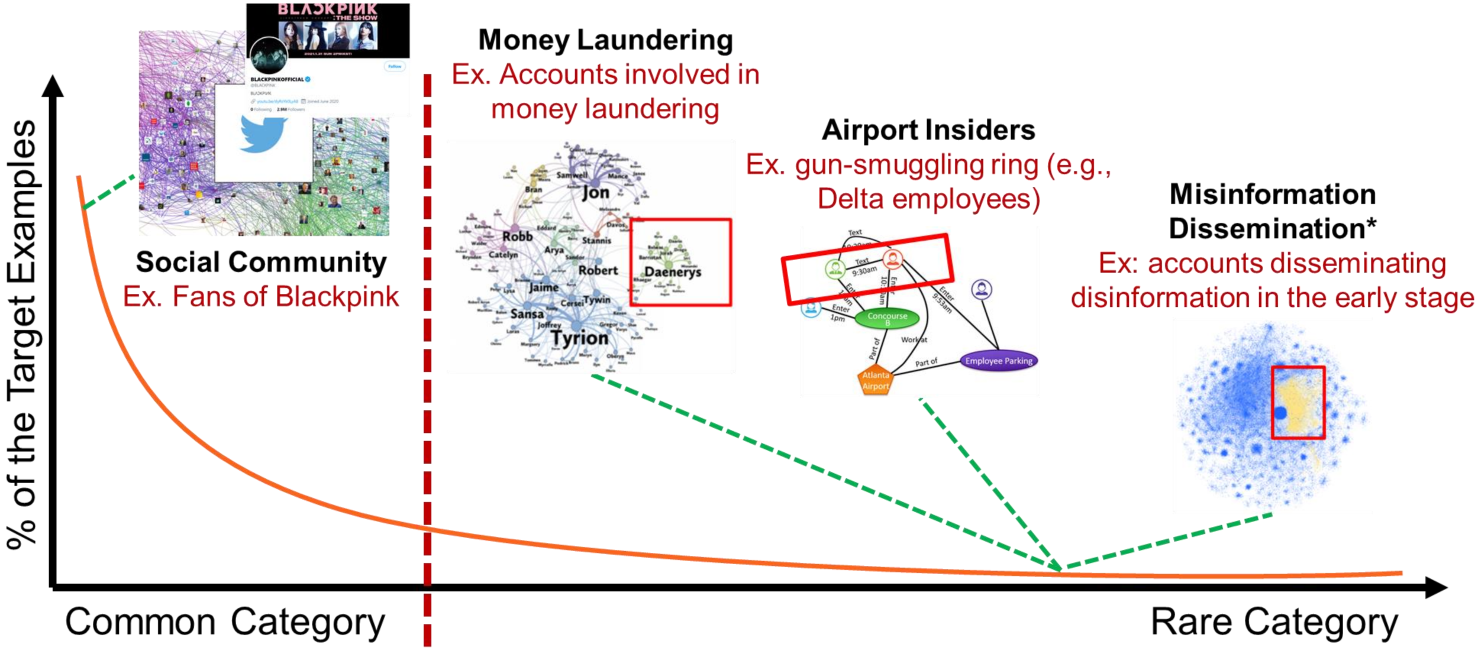
- West, Jarrod, et al. "Intelligent financial fraud detection: a comprehensive review." Computers & security. 2016.
- Hasisi, Badi, et al. "Crime and terror: Examining criminal risk factors for terrorist recidivism." Journal of Quantitative Criminology. 2020.
- Massie, Justin, et al. "America's Allies and the Decline of US Hegemony. Routledge." 2019.

Rare Category Analysis

❑ Problem definition

- Rare category analysis (RCA) refers to the problem of analyzing the underrepresented minority classes in an imbalanced data set.

❑ The nature of rare categories



Previous Studies
➢Rarity [Lin et al., 2017] ➢Non-separability [He et al., 2010] ➢Label scarcity [He et al., 2010]

- He, Jingrui, et al. "Rare category characterization." IEEE ICDM. 2010.
- Lin, Hanfei, et al. "Rclens: Interactive rare category exploration and identification." IEEE transactions on visualization and computer graphics. 2017
- Nguyen, Hung, et al. "End-to-End Continual Rare-Class Recognition with Emerging Novel Subclasses." ACM TKDD. 2020.

Rare Category Analysis – Challenges

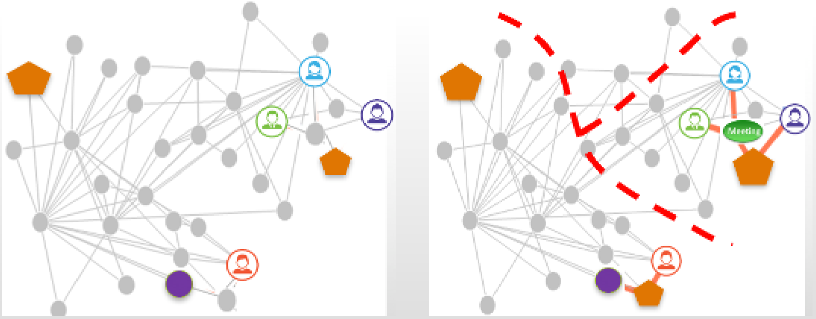
❑ Problem definition

- Rare category analysis (RCA) refers to the problem of analyzing the underrepresented minority classes in an imbalanced data set.

❑ Emerging challenges in the era of big data

Q1: How to characterize rare categories with data and task heterogeneity?

M. 1: Characterization



User Network   Local Clusters

C1. Heterogeneity

- Various information sources.

Email   Website   Social Network

- Various types of targets.

Insider threat (Individual behaviors)   Money Laundering (Group behaviors)

- Zhou, Dawei, et al. "MUVIR: Multi-View Rare Category Detection." IJCAI. 2015.
- Zhou, Dawei, et al. "A local algorithm for structure-preserving graph cut." ACM SIGKDD. 2017.
- Ji, Heng, et al. "Refining event extraction through cross-document inference." ACL. 2008.

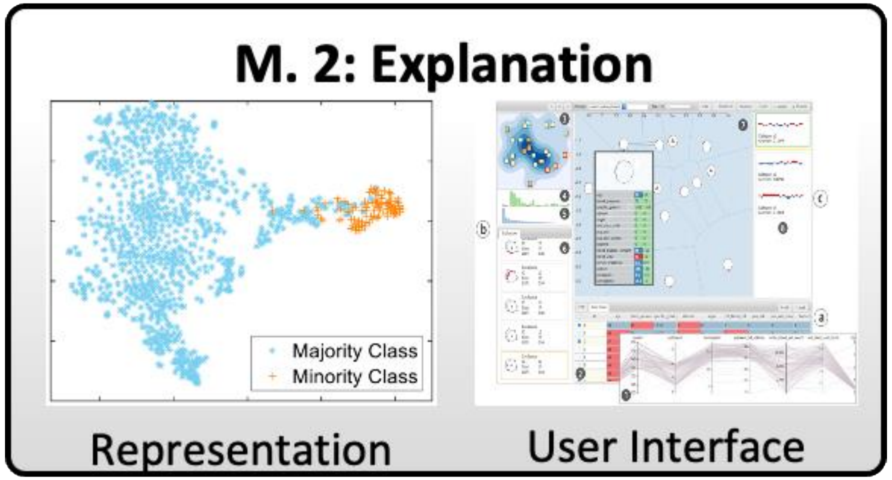
Rare Category Analysis – Challenges

❑ Problem definition

- Rare category analysis (RCA) refers to the problem of analyzing the underrepresented minority classes in an imbalanced data set.

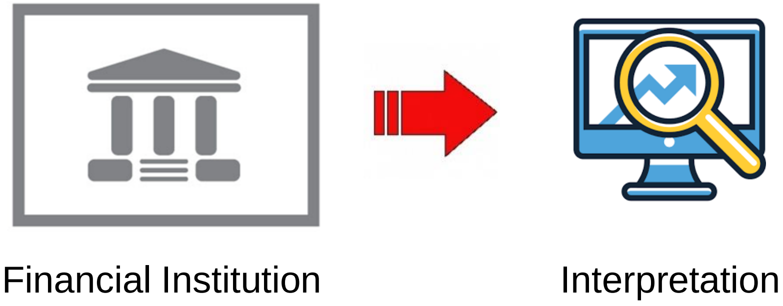
❑ Emerging challenges in the era of big data

Q2: How to interpret the prediction results and provide relevant clues for the end-users?



C2. Interpretability

- Many machine learning models remain "black-box" in nature.
- Traditional industries require models to be interpretable.



- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM. 2018.
- Pan, Jia-cheng, et al. "RCAnalyzer: visual analytics of rare categories in dynamic networks." Frontiers of Information Technology & Electronic Engineering. 2020.
- Yang, Carl, et al. "Neural Concept Map Generation for Effective Document Classification with Interpretable Structured Summarization." ACM SIGIR. 2020.

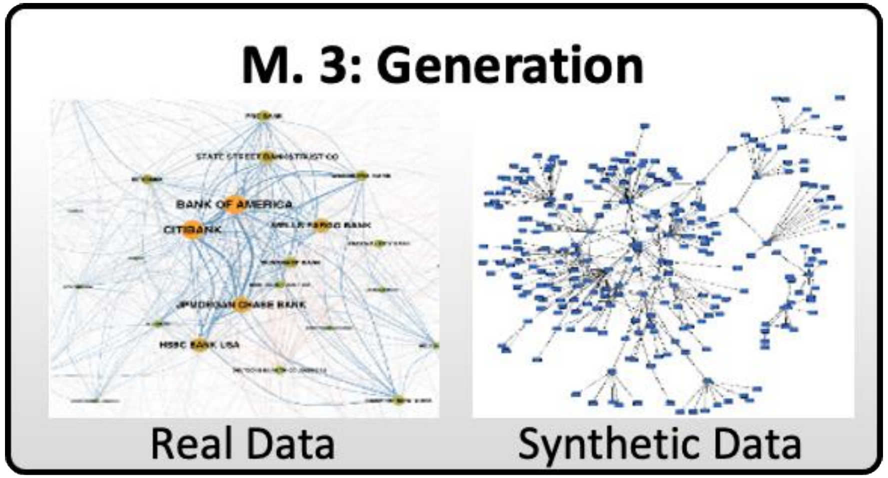
Rare Category Analysis – Challenges

❑ Problem definition

- ○ Rare category analysis (RCA) refers to the problem of analyzing the underrepresented minority classes in an imbalanced data set.

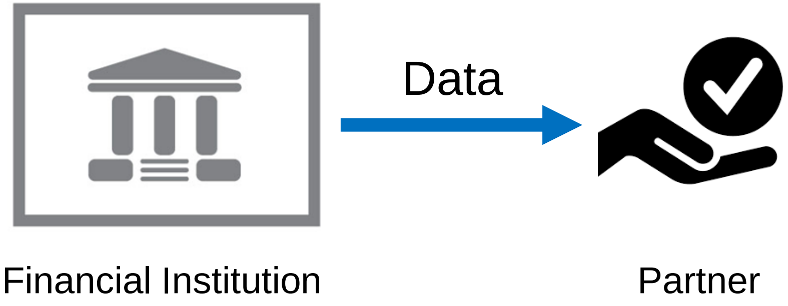
❑ Emerging challenges in the era of big data

Q3: How to alleviate the data scarcity and enable data augmentation?



C3. Data Scarcity

- ➢ Massive data containing sensitive information (e.g., sex, age).
- ➢ Data security and privacy.
- ➢ Data proprietary.



- Zhou, Dawei, et al. "A Data-Driven Graph Generative Model for Temporal Interaction Networks." ACM. 2020.
- Zhou, Dawei, et al. "Misc-GAN: A Multi-scale Generative Model for Graphs." Frontiers in Big Data. 2019.
- Akoglu, Leman, and Christos Faloutsos. "RTG: a recursive realistic graph generator using random typing." ECML PKDD. 2009.

Comparison with Imbalanced Classification

❑ Imbalanced classification

- Labeled examples from all the classes.
- Models for imbalanced classification focus on overall performance of all classes.
- **Methodology**: [Kubat & Matwin, ICML1997]; [Chawla et al, JAIR2002]; [Wu & Chang, ICML2003]; [Huang et al., CVPR2016]; …

❑ Rare category analysis

- None/one/few-shot labeled example from the rare categories.
- Models for rare category analysis put heavy emphasis on learning minority classes with a good performance.
- **Methodology**: [Fine & Mansour, COLT2006]; [Dasgupta & Hsu, ICML2008]; [Vatturi & Wong, KDD2009]; [Zhou et al., KDD2018]; …

- Branco, Paula, et al. "A survey of predictive modeling on imbalanced domains." ACM Computing Surveys. 2016.

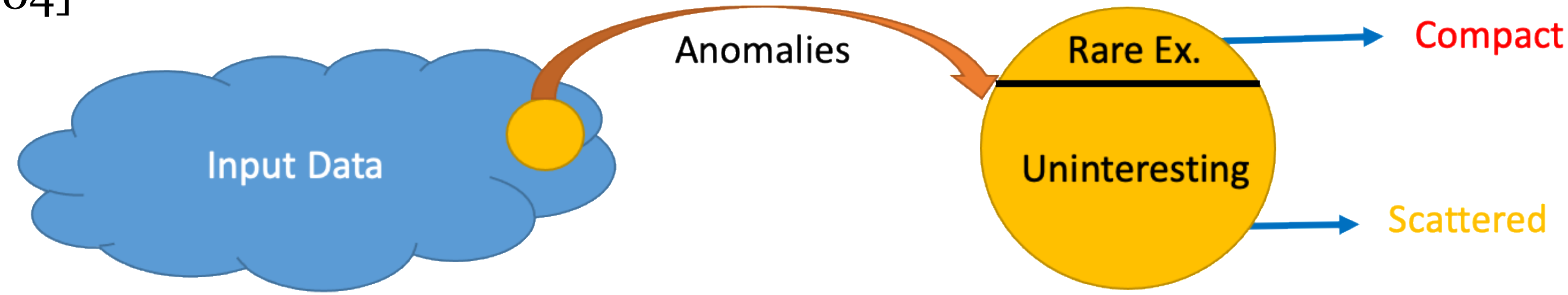
Comparison with Outlier/Anomaly Detection

❑ Outliers/anomalies

- Patterns in data that do not conform to expected behavior [Chandola, Banerjee & Kumar, 2009].
- Typically separable from normal examples and scattered in the feature space.

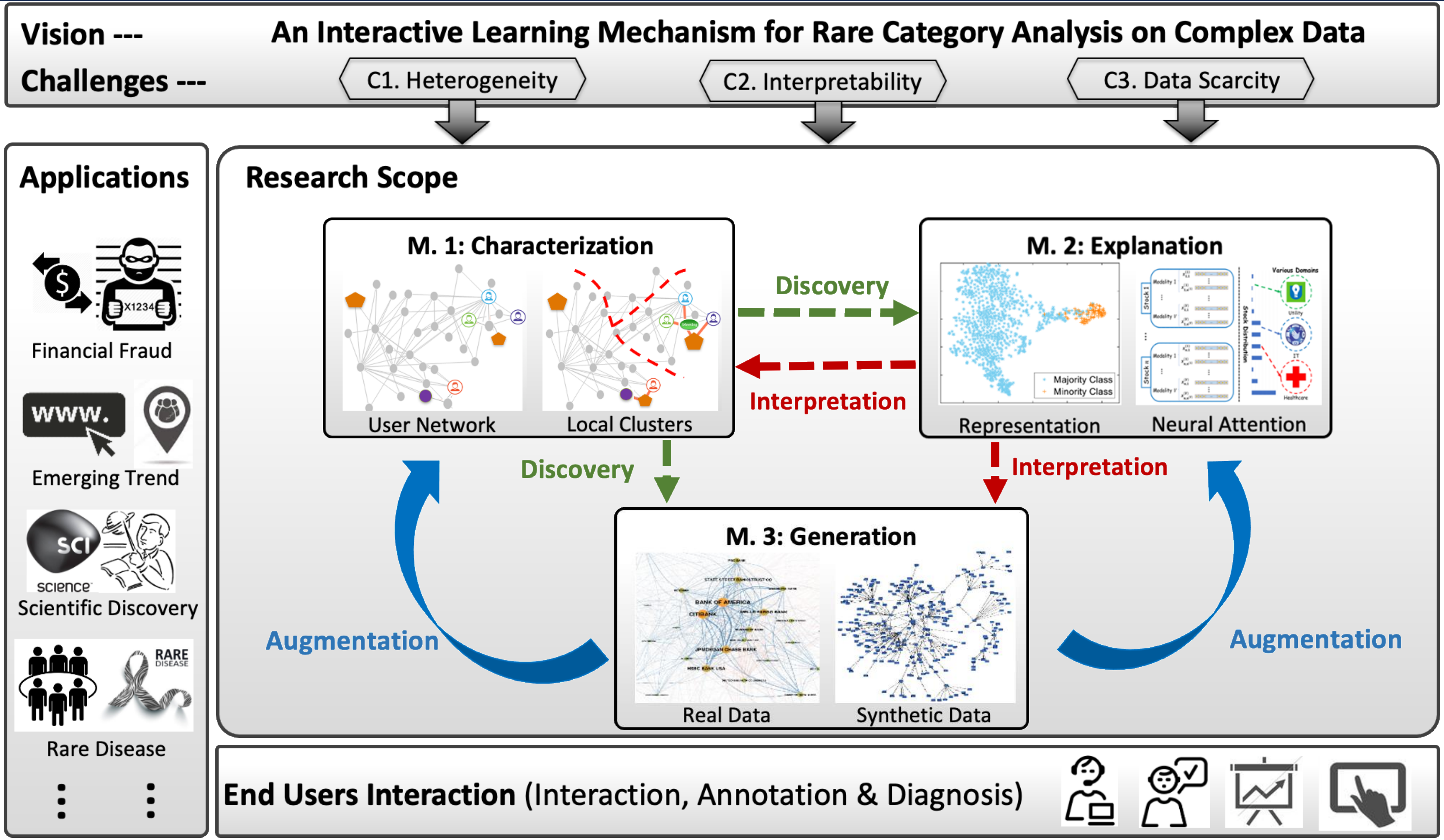
❑ Rare categories

- Compact in the feature space.
- “Most of the objects (i.e., majority classes) are well explained by current theories and .... remainder are anomalies, but 99% of these anomalies are uninteresting (i.e., boring anomalies), and only 1% of them (i.e., rare categories) are useful” [Pelleg & Moore, 2004]

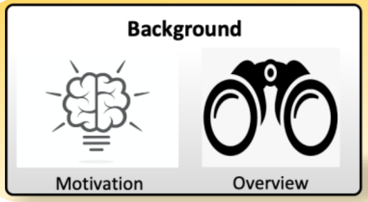


- Gupta, Manish, et al. "Outlier detection for temporal data: A survey." IEEE Transactions on Knowledge and data Engineering. 2013.
- Akoglu, Leman, et al. "Graph based anomaly detection and description: a survey." Data mining and knowledge discovery. 2015.

Overview of Rare Category Trinity

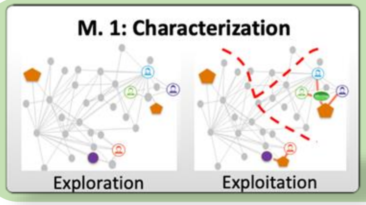


Roadmap



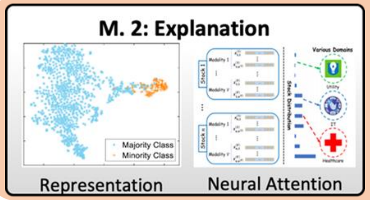
Background

- Motivation
- Research overview



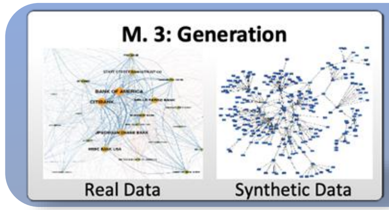
I. Rare Category Characterization

- Rare category characterization on homogeneous graphs
- Rare category characterization on heterogeneous graphs



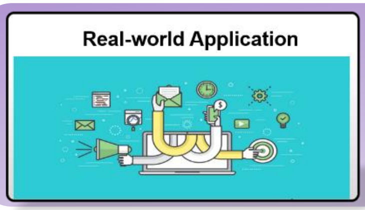
II. Rare Category Explanation

- Data insights
- Model insights



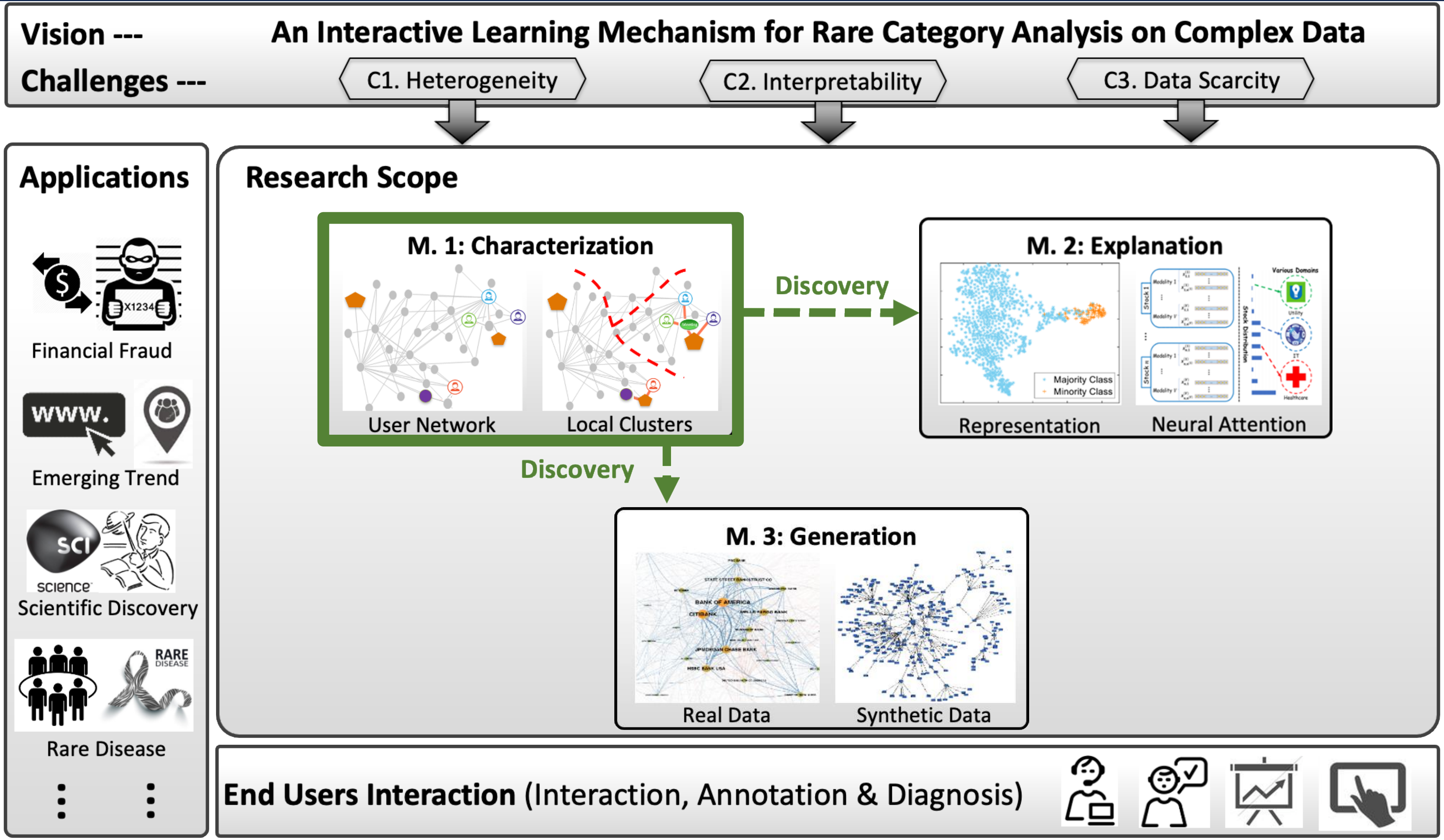
III. Rare Category Generation

- Unsupervised rare category generation
- Supervised rare category generation



IV. Real-world Application

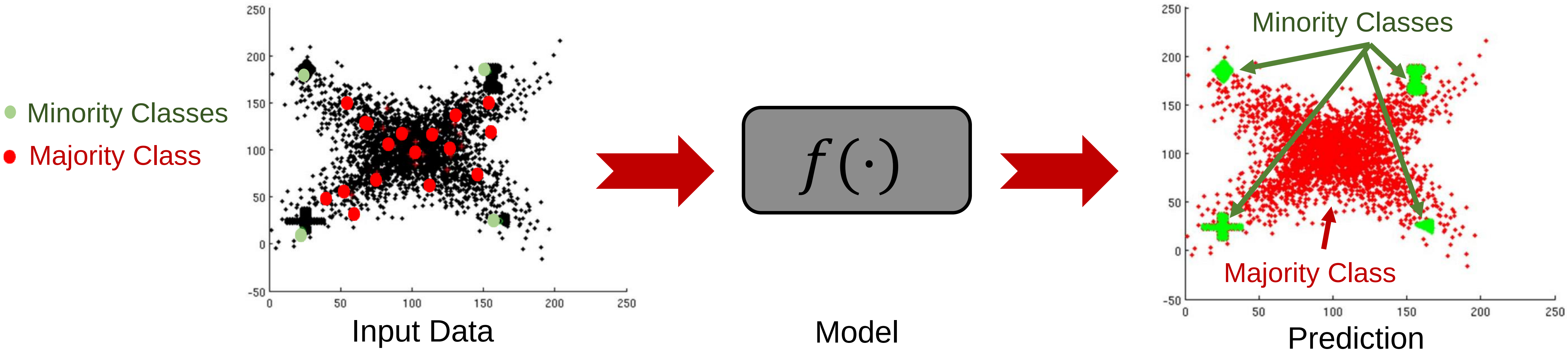
Rare Category Characterization



Rare Category Characterization

Problem 1.1. Rare category characterization

- **Given**: (1) A data set $D = \{x_1, \ldots, x_n\}$, (2) a handful of labeled examples $L = \{x_1, \ldots, x_{|L|}\}$ from both minority classes and majority classes, (3) unlabeled examples U.
- **Find**: A compact representation that characterizes minority class examples.



- He, Jingrui, et al. "Rare category characterization." IEEE ICDM. 2010.

Rare Category Characterization

Part I: Rare Category Characterization

Homogeneous Graphs

- Node-level characterization
- Motif-level characterization

Heterogeneous Graphs

- Bipartite graphs
- Heterogeneous graphs

Node-Level Characterization on Static Graphs

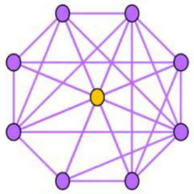
- **Problem 1.2. Rare category characterization on weighted graphs**
  - **Given**: a weighted graph $G = (V, E)$.
  - **Find**: a list of potential rare examples, which are strange, abnormal, extremely rare.

- **Challenges**
  - How to extract relevant features?
  - How to spot the nodes that are anomalous (rare examples) ?

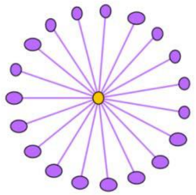
OddBall Algorithm

Feature extraction for egonet

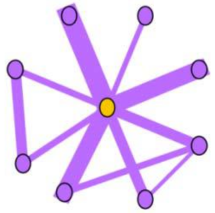
- N_i: number of neighbors (degree) of egonet i



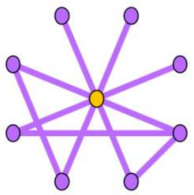
- E_i: number of edges in egonet i

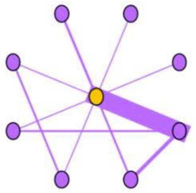


- W_i: total weight of egonet i



- $\lambda_{W,i}$: principal eigenvalue of the weighted adjacency matrix of egonet i

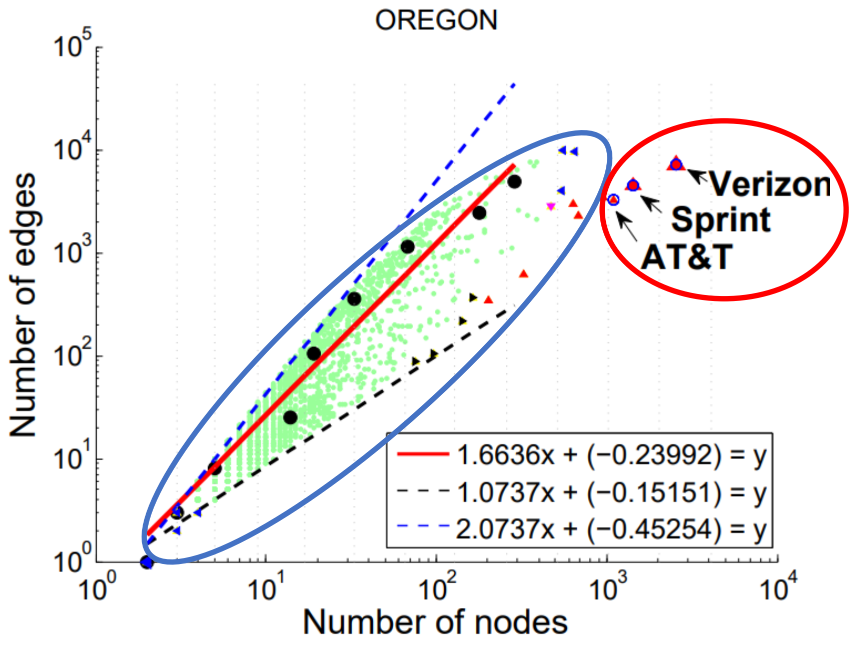




- Akoglu, Leman, et al. "OddBall: Spotting Anomalies in Weighted Graphs." PAKDD, 2010.

OddBall Algorithm

- **Proposed method**
  - For each node,
    - Extract "ego-net" (=1-step neighborhood)
    - Extract features (#edges, total weight, etc.)
      - features that could yield "laws"
      - features fast to compute and interpret
  - Characterize regular patterns
    - Examples from majority classes
    - Compute the distribution of regular patterns
  - Detect irregular patterns
    - Examples from minority classes
    - Intuition: far from the fitting line → anomaly

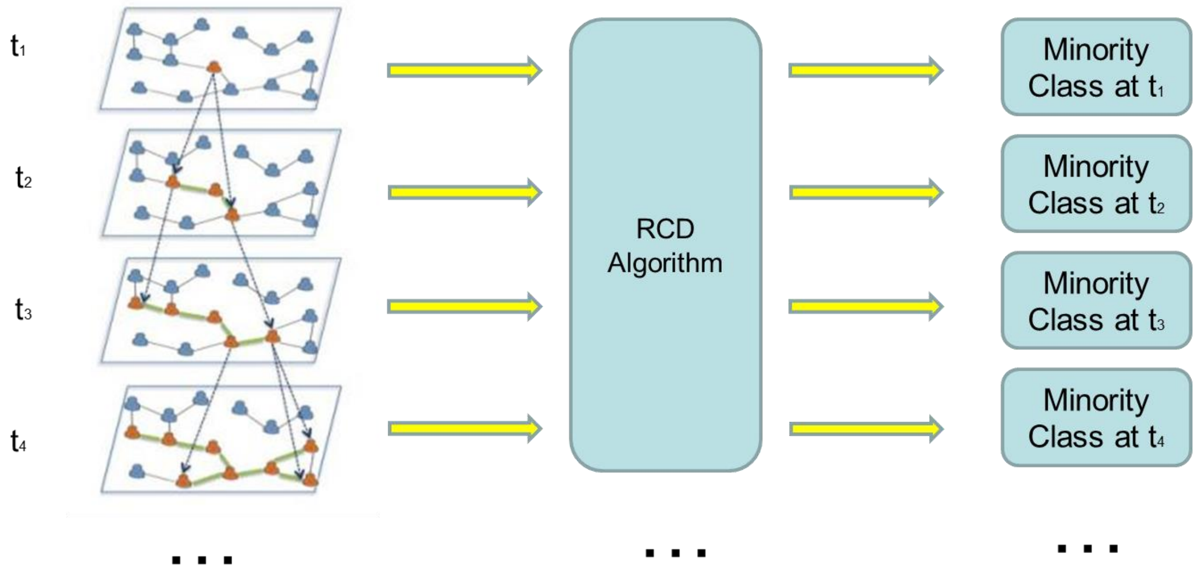


- Akoglu, Leman, et al. "OddBall: Spotting Anomalies in Weighted Graphs." PAKDD, 2010.

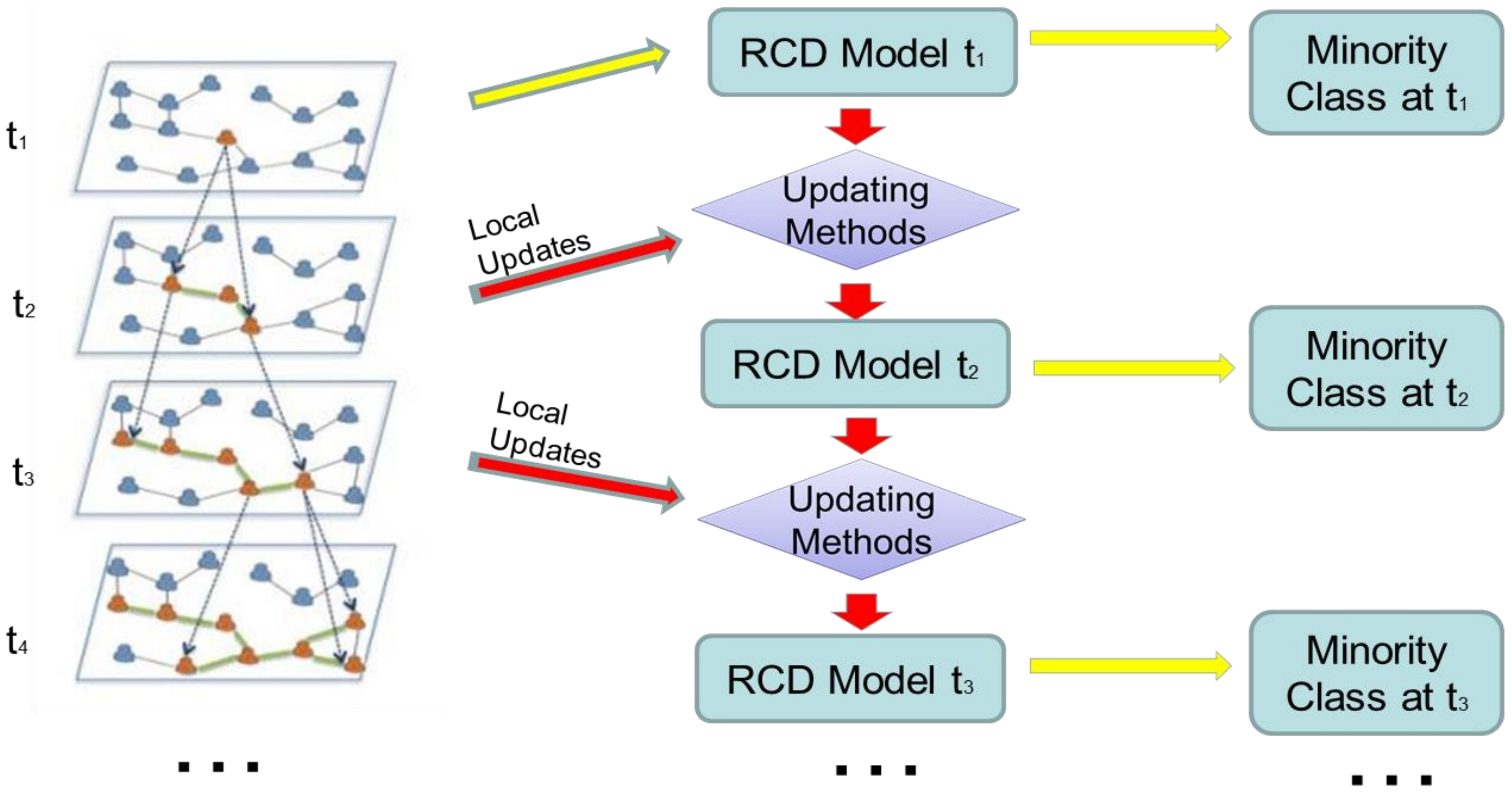
Node-Level Characterization on Dynamic Graphs

- **Problem 1.3. Rare category characterization on Time-evolving Graphs**
  - **Given**: a time-evolving graphs $\tilde{G} = \{G_1, G_2, \dots, G_T\}$
  - **Find**: examples from each minority class at each snapshot G_t , t=1, …, T.

- **Challenges**
  - C1: Graphs are evolving over time
    - New nodes/edges show up/die out
    - Edge weights change
  - C2: High computational cost
    - Space complexity
    - Time complexity



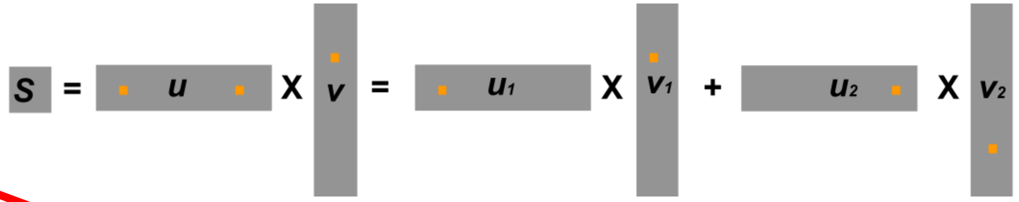
BIRD Algorithm

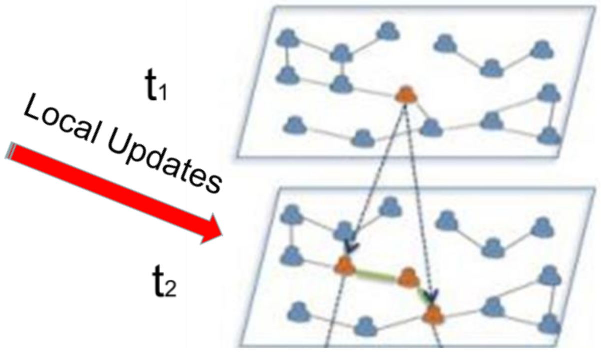


- Zhou, Dawei, et al. “Rare Category Detection on Time-Evolving Graphs.” ICDM, 2015.

❑ Proposed methods

- Extract the updating matrix ***S*** from the last timestamp.
- Iteratively update ***A*** with each updated edge based on Sherman-Morrison formula.





Previous timestamp

$$A^{(t)} = A^{(t-1)} + \alpha \frac{A^{(t-1)} u v^T A^{(t-1)}}{I + v^T A^{(t-1)} u}$$

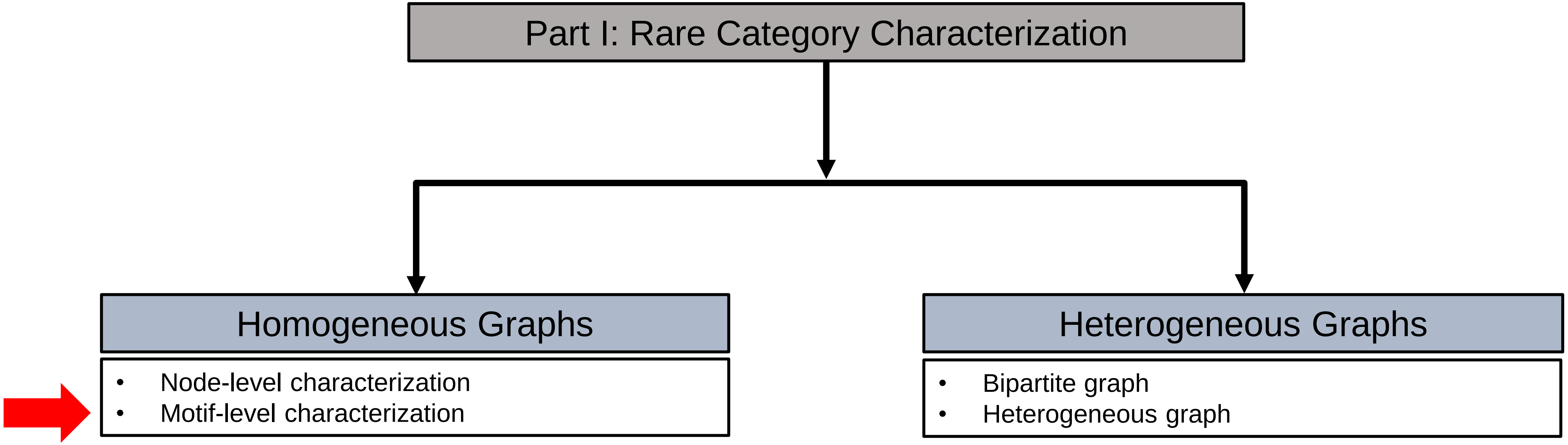
Graph Changes

- Compute K-NN matrix at the current timestamp.
  - ➢ Only update the rows in which the order of elements is changed in the evolved graph.

- Compute local density changes of each node.
  - Largest local density changes → potential rare category examples.

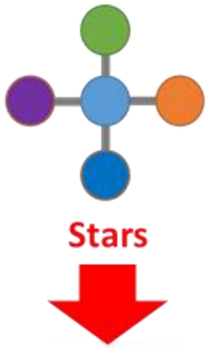
- Zhou, Dawei, et al. "Rare Category Detection on Time-Evolving Graphs." ICDM, 2015.

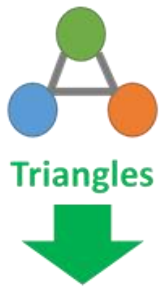
Rare Category Characterization

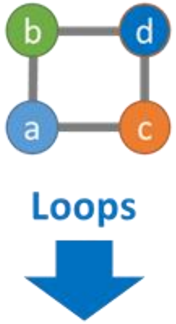


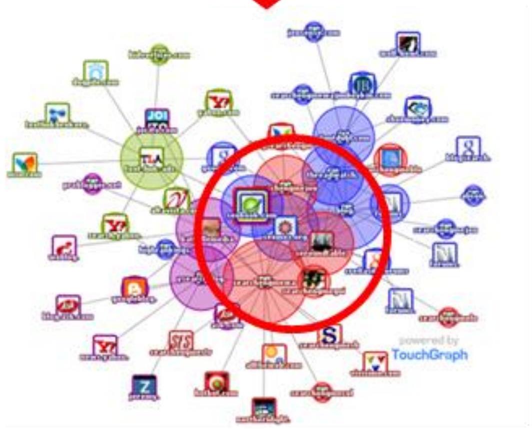
Motif-Level Characterization

- Higher-order connectivity patterns, or network motifs, mediate complex networks.



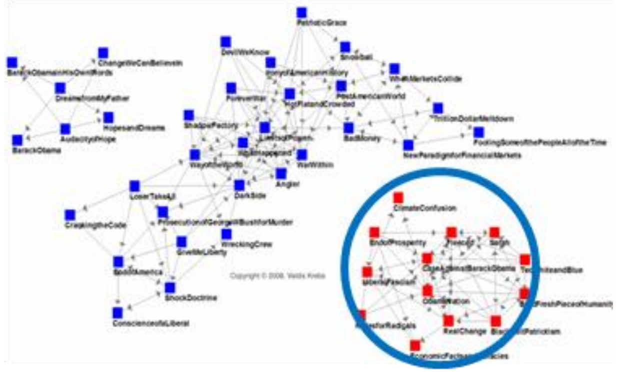






Internet

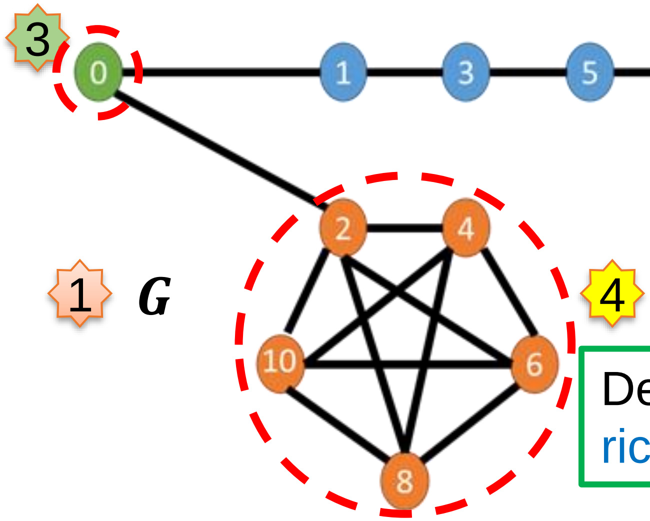
Social Network



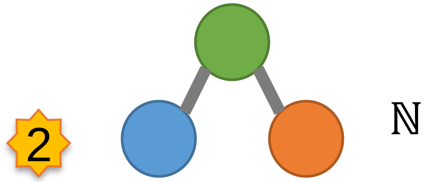
Online Transaction

- West, Jarrod, et al. "Intelligent financial fraud detection: a comprehensive review." Computers & security. 2016.
- Kim, Jooho, and Makarand Hastak. "Social network analysis: Characteristics of online social networks after a disaster." International Journal of Information Management. 2018.

- **Problem 1.4. Structure-preserving local graph cut**
  - **Given:** Graph $G=(V, E)$, a user-defined motif $\mathbb{N}$, an initial node v.
  - **Find:** A structure-rich dense subgraph that (1) includes or near the initial node v and (2) minimizes the motif conductance w.r.t. the user-defined $\mathbb{N}$.



Dense subgraph with rich wedges

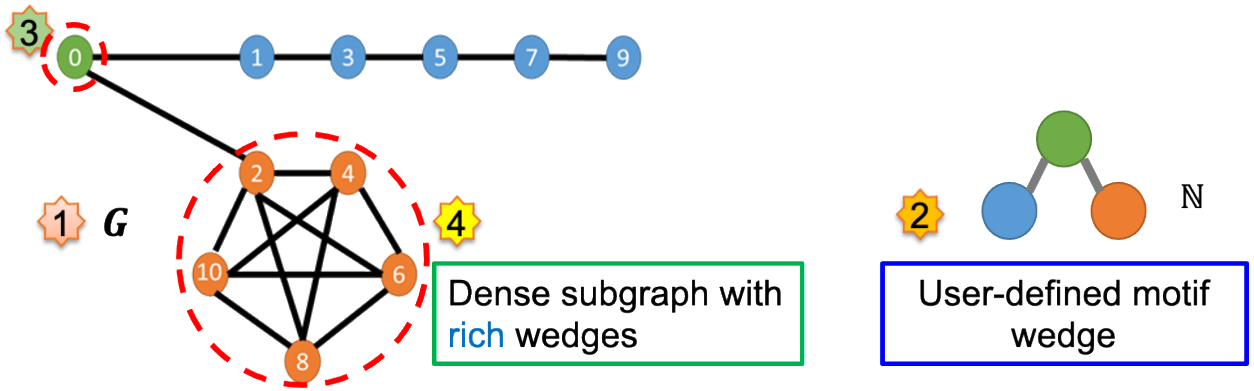


User-defined motif wedge

- Zhou, Dawei, et al. "A local algorithm for structure-preserving graph cut." ACM SIGKDD. 2017.

Problem

- **Problem 1.4 is NP-hard**
  - Any generalization of the conductance minimization problem is NP-hard [Wagner and Wagner, 1993].

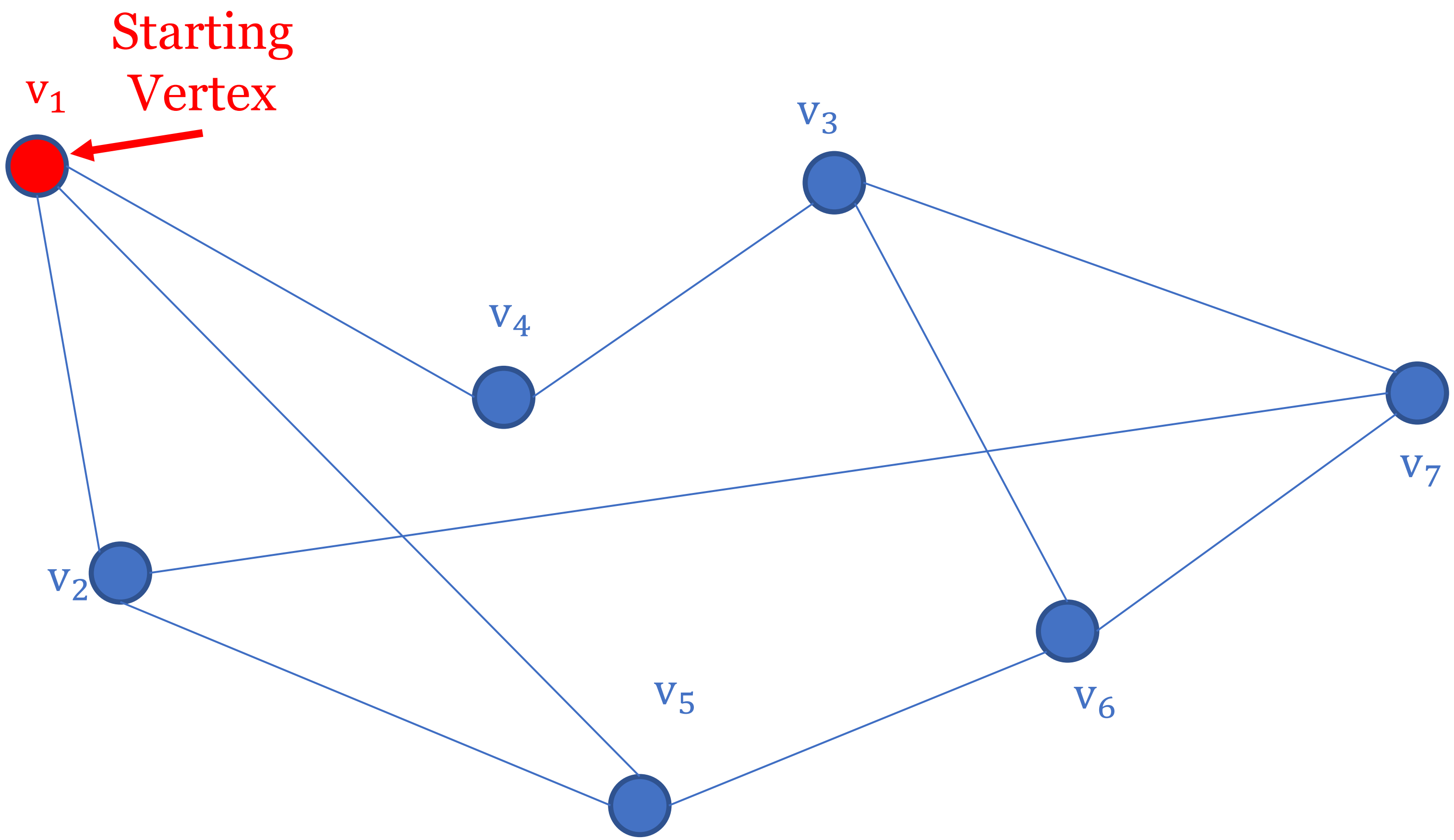


- **An approximation solution**
  - **HOSPLOC - High-Order Structure-Preserving LOcal Graph Cut.**
  - A random walk based method.

- Wagner, Dorothea, and Frank Wagner. "Between min cut and graph bisection." MFCS, 1993.
- Spielman, Daniel A., and Shang-Hua Teng. "A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning." SIAM Journal on computing. 2013.

Graph Random Walk

- A random walk on a graph is a process that begins at some vertex, and at each time step moves to another vertex.

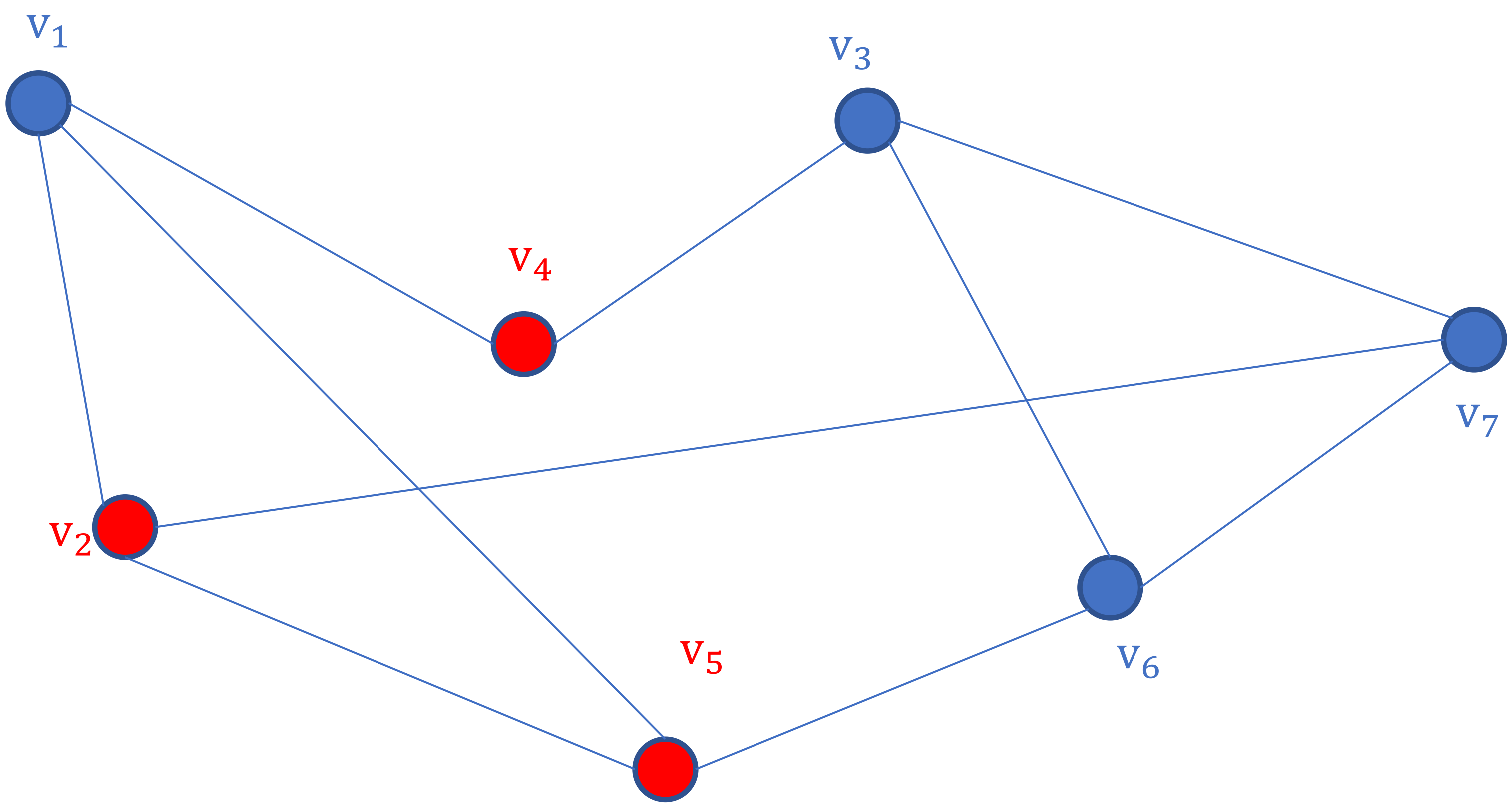


Current Random walk sequence:
$W = \{v_1\}$

- Tong, Hanghang, Christos Faloutsos, and Jia-Yu Pan. "Fast random walk with restart and its applications." *Sixth international conference on data mining (ICDM'06)*. IEEE, 2006.

Graph Random Walk

- A random walk on a graph is a process that begins at some vertex, and at each time step moves to another vertex.

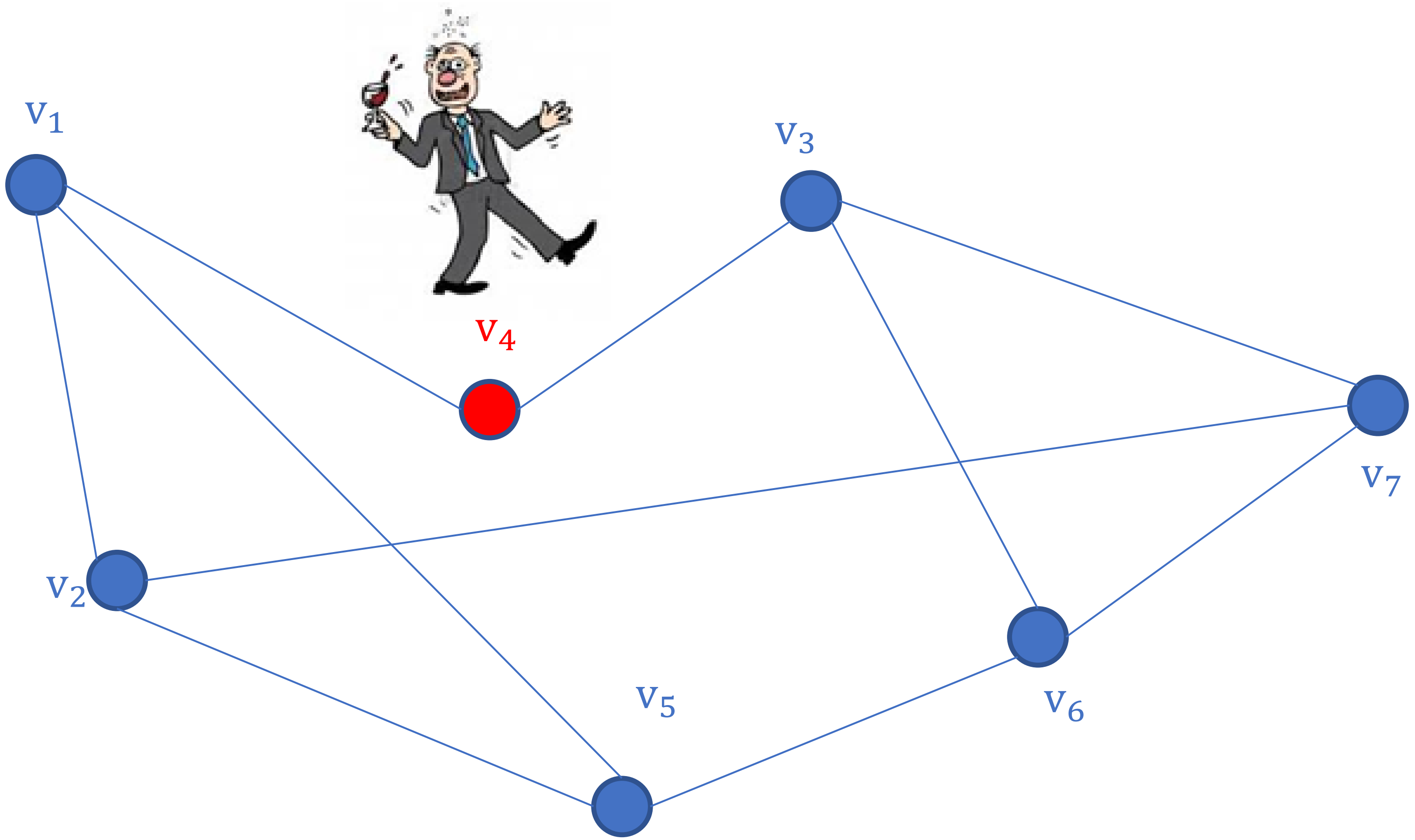


Current Random walk sequence:
$W = \{v_1\}$

- Tong, Hanghang, Christos Faloutsos, and Jia-Yu Pan. "Fast random walk with restart and its applications." *Sixth international conference on data mining (ICDM'06)*. IEEE, 2006.

Graph Random Walk

- A random walk on a graph is a process that begins at some vertex, and at each time step moves to another vertex.

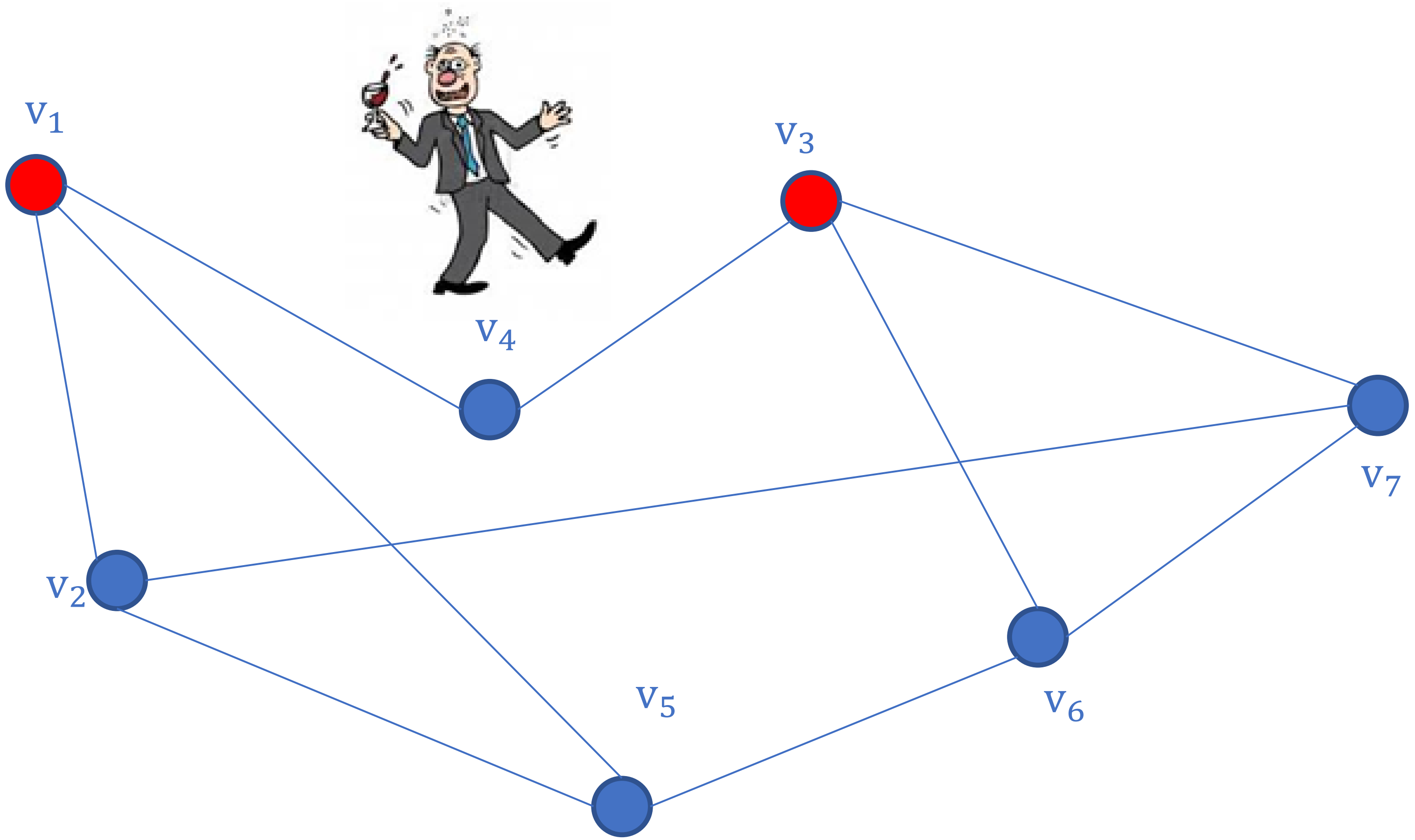


Current Random walk sequence:
$W = \{v_1, v_4\}$

- Tong, Hanghang, Christos Faloutsos, and Jia-Yu Pan. "Fast random walk with restart and its applications." *Sixth international conference on data mining (ICDM'06)*. IEEE, 2006.

Graph Random Walk

- A random walk on a graph is a process that begins at some vertex, and at each time step moves to another vertex.

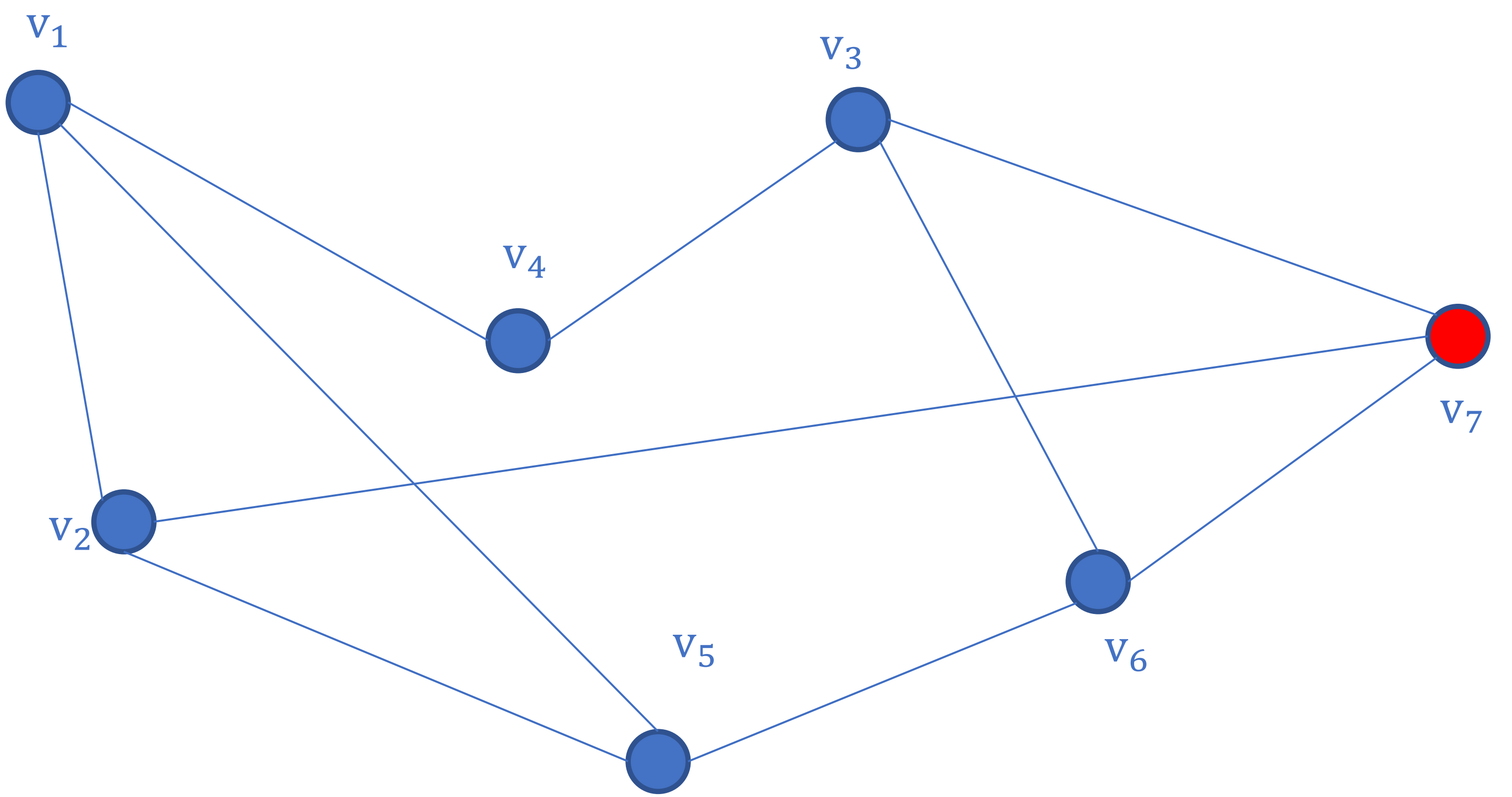


Current Random walk sequence:
$W = \{v_1, v_4\}$

- Tong, Hanghang, Christos Faloutsos, and Jia-Yu Pan. "Fast random walk with restart and its applications." *Sixth international conference on data mining (ICDM'06)*. IEEE, 2006.

Graph Random Walk

- ❑ A random walk on a graph is a process that begins at some vertex, and at each time step moves to another vertex.



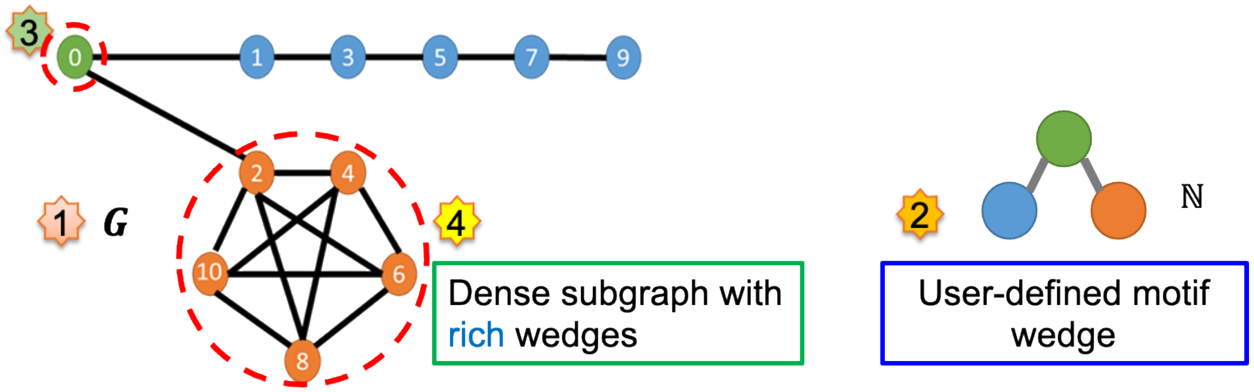
k-length Random walk sequence:

$$W = \underbrace{\{v_1, v_4, v_3, \dots, v_7\}}_{k}$$

- Tong, Hanghang, Christos Faloutsos, and Jia-Yu Pan. "Fast random walk with restart and its applications." *Sixth international conference on data mining (ICDM'06)*. IEEE, 2006.

Problem

- **Problem 1.4 is NP-hard**
  - Any generalization of the conductance minimization problem is NP-hard [Wagner and Wagner, 1993].



- **An approximation solution**
  - **HOSPLOC - H**igh-**O**rder **S**tructure-**P**reserving **LO**cal Graph **C**ut.
  - A random walk based method.
  - Space and time complexity could be greatly reduced.

- Wagner, Dorothea, and Frank Wagner. "Between min cut and graph bisection." MFCS, 1993.
- Spielman, Daniel A., and Shang-Hua Teng. "A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning." SIAM Journal on computing. 2013.

HOSPLOC Algorithm

❑ Algorithm overview

- **Step 1.** Modeling high-order motifs by constructing adjacency tensor of the original graph.
- **Step 2.** Exploring neighborhood context around initial node via truncated high-order random walk.
- **Step 3.** Conducting structure-preserving graph cut via sweep cut procedure on the stationary distribution vectors of high-order random walk.

❑ Properties

- **Quality guarantee**
  - Finding a near-optimal local cluster with minimum motif conductance.
- **Scalability guarantee**
  - Procedure stops upon finding a good cluster, no need to explore the rest of graph.
  - Algorithm runs in polylogarithmic time w.r.t. the number of edges of original graph.

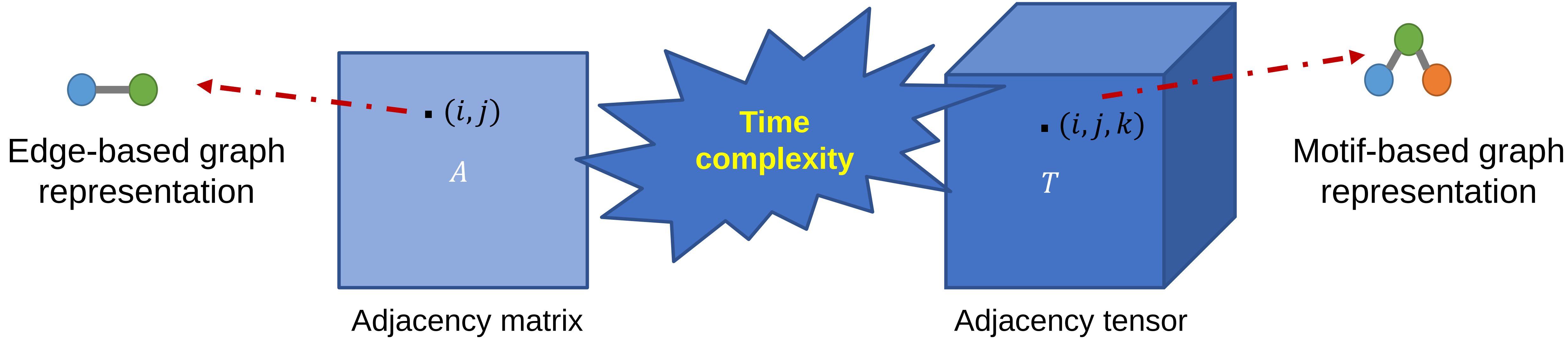
- Wagner, Dorothea, and Frank Wagner. "Between min cut and graph bisection." MFCS. 1993.
- Spielman, Daniel A., and Shang-Hua Teng. "A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning." SIAM Journal on computing. 2013.

HOSPLOC – Step I: Modeling Network Motifs

❑ Adjacency tensor

Definition 1.1. Given a graph G=(V,E), the k^{th}-order network motif ℕ on G could be represented in a k-dimensional adjacency tensor T as follows

$$T(i_1, i_2, \dots, i_k) = \begin{cases} 1 & (i_1, i_2, \dots, i_k) \in V \text{ and form } \mathbb{N} \\ 0 & \text{Otherwise} \end{cases}$$



- Zhou, Dawei, et al. "A local algorithm for structure-preserving graph cut." ACM SIGKDD. 2017.

HOSPLOC – Step II: Exploring Neighborhood

❑ Truncated high-order random walk (THRW)

- Exploring k^{th}–order network structure $\mathbb{N}$.

Transition tensor

$$P(i_1, i_2, \dots, i_k) = \frac{T(i_1, i_2, \dots, i_k)}{\sum_{i_1=1}^{n} T(i_1, i_2, \dots, i_k)}$$

- Stationary distribution of HRW with "rank-1" assumption [Li and NG, 2013] .

HRW stationary distribution

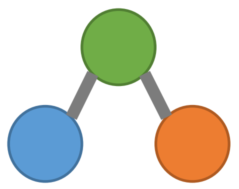
$$q(i_1) = \sum_{(i_2,\dots,i_k)} P(i_1, i_2, \dots, i_k)\, q(i_2) \dots q(i_k) \quad O(n^k)!$$

- Rounding small values of q to 0 to reduce time complexity.

Cluster indicator vector

$$[q]_\epsilon(u) = \begin{cases} q(u) & \text{If } q(u) > d(u)\epsilon \\ 0 & \text{Otherwise} \end{cases} \quad \text{Sparse computation!}$$

Ex:

$\mathbb{N}$:  $k = 3$

2nd-order HRW

$\boldsymbol{P} \in \mathbb{R}^{n \times n \times n}$

$\boldsymbol{q} \in \mathbb{R}^{n}$

$[\boldsymbol{q}]_\epsilon \in \mathbb{R}^{n}$

- Li, Wen, and Michael K. Ng. "On the limiting probability distribution of a transition probability tensor." Linear and Multilinear Algebra. 2014.
- Benson, Austin R., et al. "The spacey random walk: A stochastic process for higher-order data." SIAM Review. 2017.

HOSPLOC – Step III: Conducting Sweep Cut

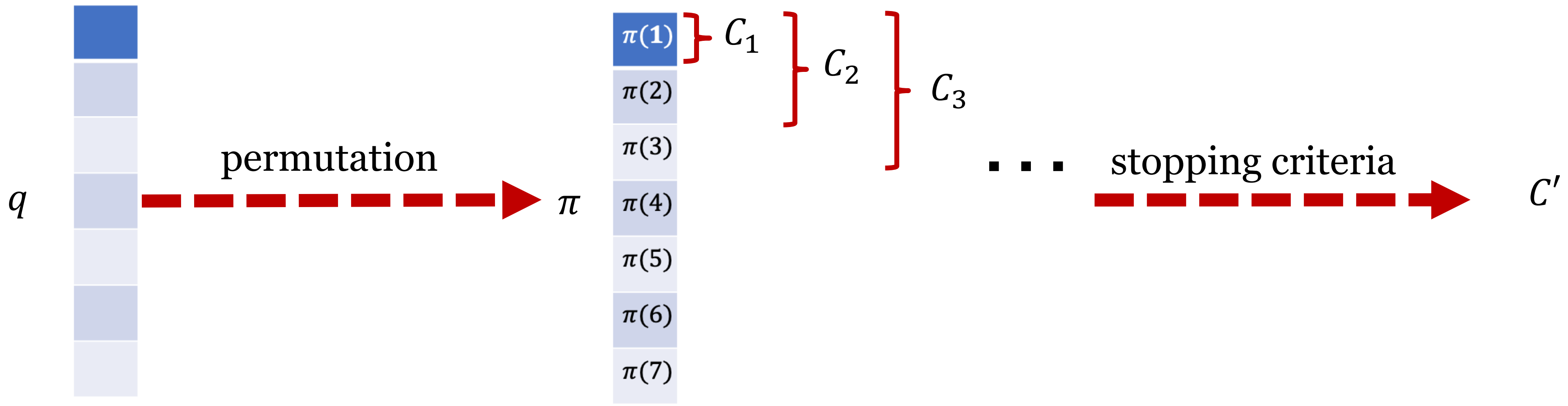
❑ Sweep cut procedure

- Obtaining high-order random walk distribution by performing THRW
- Computing the permutation π of vector q in the way that:

$$\frac{q(\pi(1))}{d(\pi(1))} \geq \frac{q(\pi(2))}{d(\pi(2))} \geq \cdots \geq \frac{q(\pi(n))}{\boxed{d(\pi(n))}}$$

Node Degree

- Iteratively checking the potential cuts $C_1, C_2, \dots, C_{n-1}$, where $C_i = \{\pi(1), \dots, \pi(i)\}$, and identify the first cut C' that satisfies the user-defined stopping criteria.



- Zhou, Dawei, et al. "A local algorithm for structure-preserving graph cut." ACM SIGKDD. 2017.
- Spielman, Daniel A., et al. "A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning." SIAM Journal on computing. 2013.

MAPPR Algorithm

- **Motif-based approximate personalized PageRank algorithm (MAPPR)**
  - **Step 1.** Create a weighted graph with weighted edge conductance that equals to the motif conductance in the original graph.

$$w(i, j) = \text{\#motif instances containing nodes } i \text{ and } j.$$

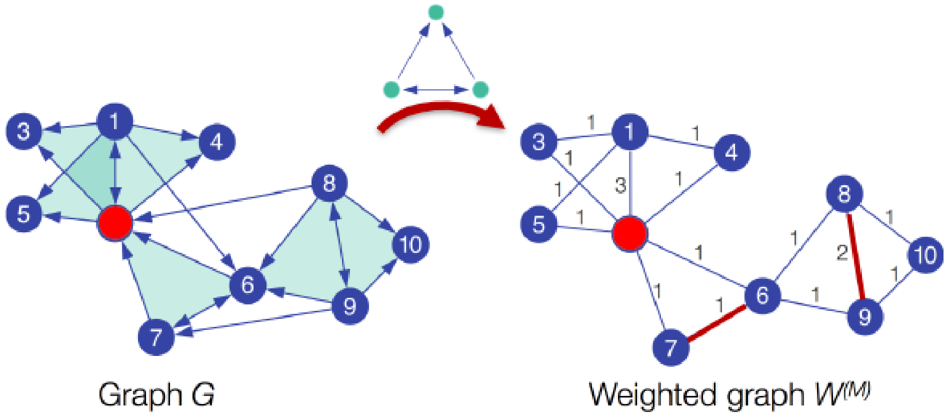


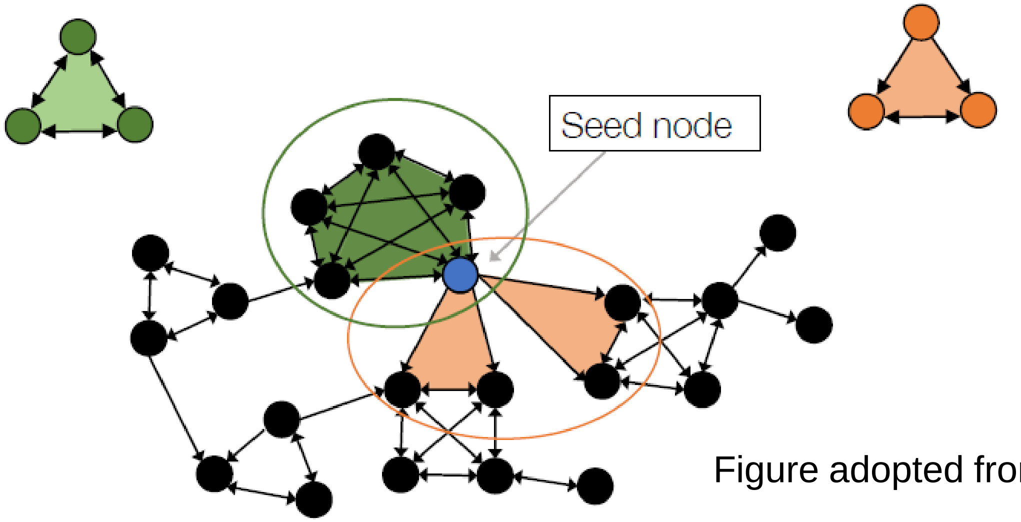
Figure adopted from [Yin et al., KDD 2017]

The motif conductance (approximately) equals the weighted edge conductance in this weighted graph [Benson et at.,16].

- Yin, Hao, et al. "Local higher-order graph clustering." ACM SIGKDD, 2017.
- Benson, Austin, et al. "Higher-order organization of complex networks." Science, 2016.

MAPPR Algorithm

- **Motif-based approximate personalized PageRank algorithm (MAPPR)**
  - **Step 1.** Create a weighted graph with weighted edge conductance equals the motif conductance in the original graph.
  - **Step 2.** Compute an approximate PPR vector for this weighted graph.
  - **Step 3.** Find a cluster of minimal weighted edge conductance.



Different motifs give different local clustering results!

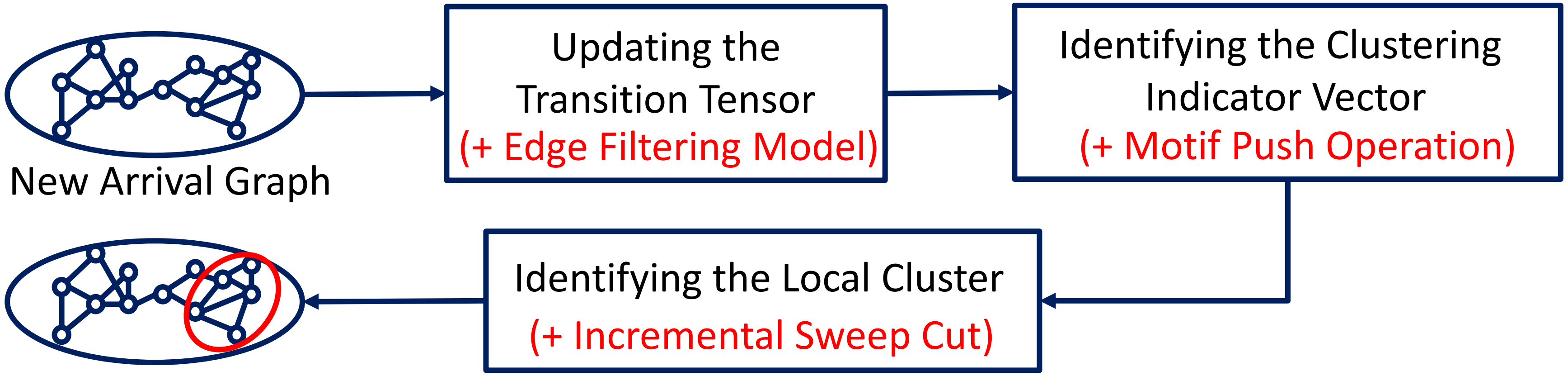
Figure adopted from [Yin et al., KDD 2017]

- Yin, Hao, et al. "Local higher-order graph clustering." ACM SIGKDD 2017.

L-MEGA Algorithm

❑ Local motif clustering on time-evolving graphs (L-MEGA)

- **Goal**: Identifying the evolution pattern of the local motif cluster effectively and efficiently.
- **Key idea**
  - ➢ Edge filtering model: Identifying "far-away" edge (v_1, v_2), filter it out before updating the transition tensor, and save it for the future.
  - ➢ Motif push operation: Tracking multilinear PageRank $\boldsymbol{q}^{(t+1)}$ from $\boldsymbol{q}^{(t)}$ via incremental updating strategies.
  - ➢ Incremental sweep cut: Identifying shared sequence between two consecutive time permutations and start from the first different entry.



- Fu, Dongqi, et al. "Local higher-order graph clustering." ACM SIGKDD 2020.

Rare Category Characterization

Part I: Rare Category Characterization

Homogeneous Graphs

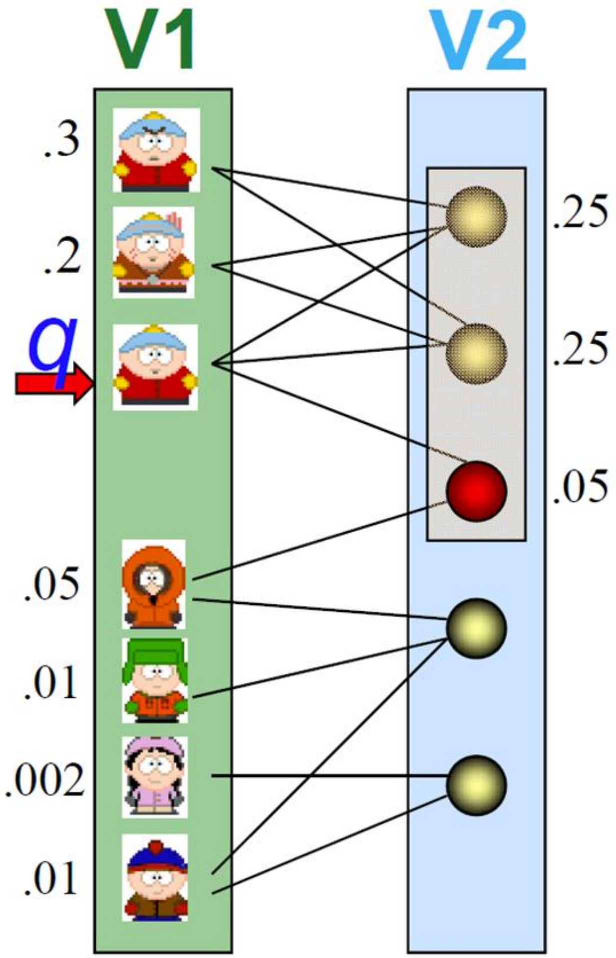
- Node-level characterization
- Motif-level characterization

Heterogeneous Graphs

- Bipartite graph
- Heterogeneous graph

Rare Category in Bipartite Graphs

- **Problem 1.6. Rare category characterization in Bipartite graphs**
  - Neighborhood formation (NF)
    - Given a query node q in $\boldsymbol{V1}$, what are the relevance scores of all the nodes in $\boldsymbol{V1}$ to q ?
    - **Ex**: Similar authors in publication networks
  - Anomaly detection (AD)
    - Given a query node q in $\boldsymbol{V1}$, what are the normality scores for nodes in $\boldsymbol{V2}$ that link to q ?
    - **Ex**: Unusual papers in publication networks

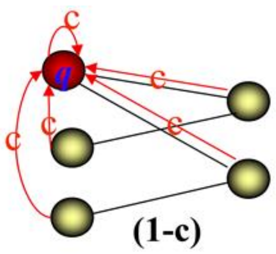


- Sun, Jimeng, et al. "Neighborhood formation and anomaly detection in bipartite graphs." ICDM, 2005.

Rare Category in Bipartite Graphs

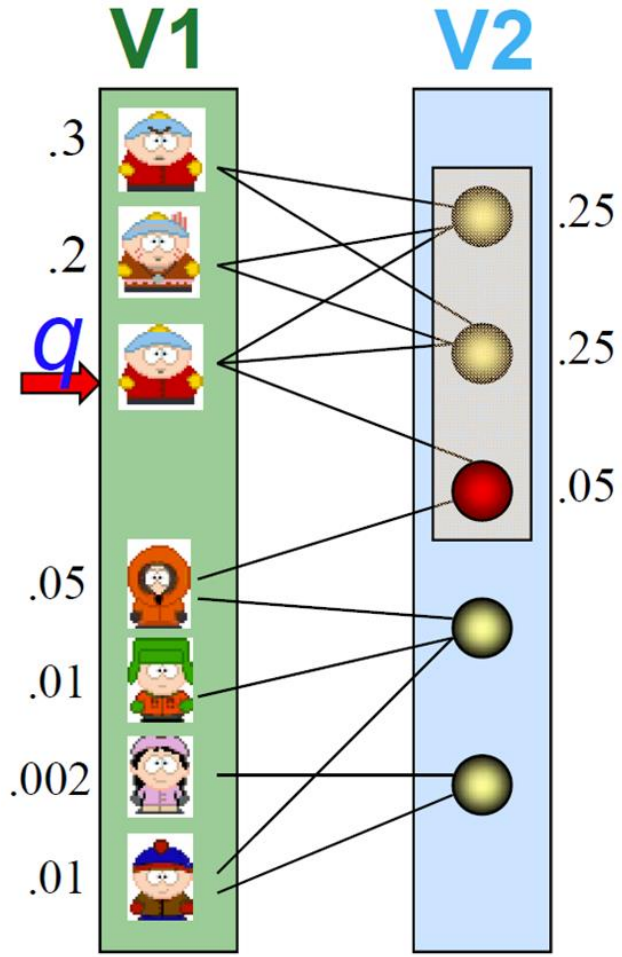
❑ Main idea for neighborhood formation

- Conduct random-walk-with-restart from q
- Compute steady-state V_1 as neighborhood relevance
  - ➢ Construct transition matrix P
  - ➢ Fly-back probability c to q



  - ➢ Solve for steady state

$$v^{(t+1)} = Pv^{(t)} + cq$$



- Sun, Jimeng, et al. "Neighborhood formation and anomaly detection in bipartite graphs." ICDM, 2005.

Rare Category in Bipartite Graphs

❑ Main idea for anomaly detection

- Pairwise "normality" scores of neighbors t
- Function of (e.g., avg) pair-wise scores
  - ➢ Find set S of nodes connected to t
  - ➢ Compute $|S|*|S|$ normality matrix R
    - Asymmetric, diagonal reset to 0
  - ➢ Apply score function $f(R)$
    - Ex: $f(R) = mean(R)$

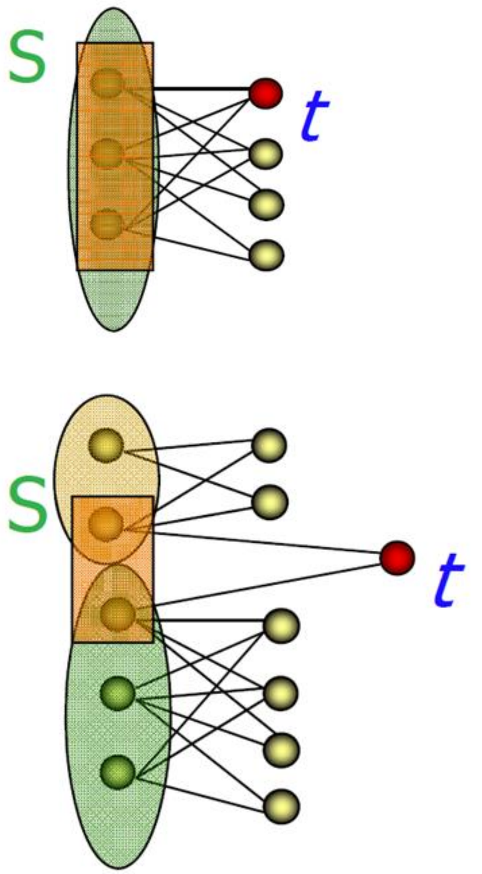


Figure adopted from [Sun et al., ICDM 2005]

- Sun, Jimeng, et al. "Neighborhood formation and anomaly detection in bipartite graphs." ICDM, 2005.

Rare Category Characterization

Part I: Rare Category Characterization

Homogeneous Graphs

- Node-level characterization
- Motif-level characterization

Heterogeneous Graphs

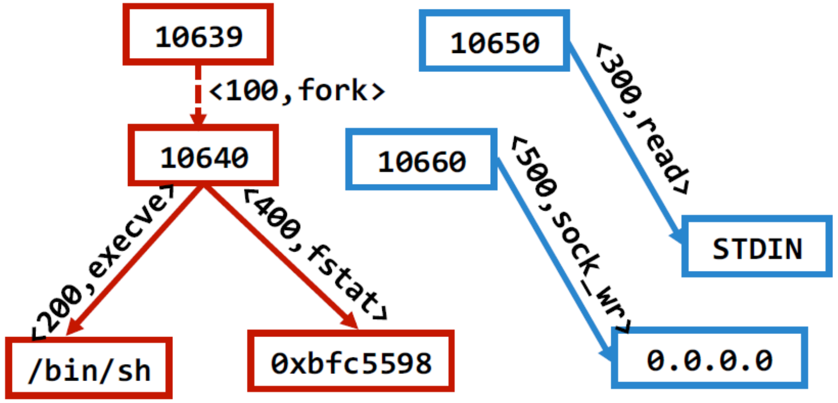
- Bipartite graph
- Heterogeneous graph

Rare Category in Heterogeneous Graphs

❑ Problem 1.7. Rare category characterization in dynamic heterogeneous graphs

- **Given:** A stream of heterogeneous graphs $G^{(t)} = (V, E, T)$ containing different types T of nodes V and edges E.
- **Find:** Rare categories in real-time while consuming bounded memory.

Example of two information flow graphs based on system logs



❑ Challenges

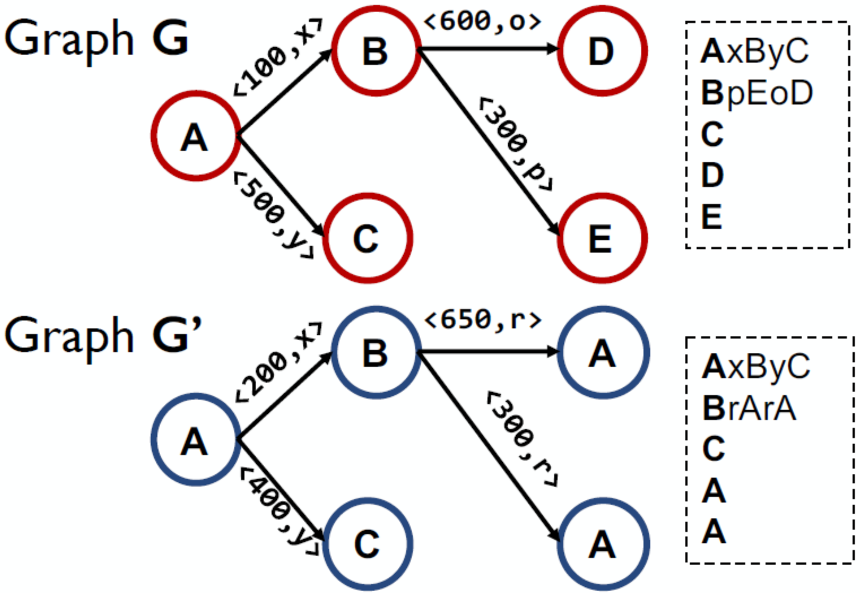
- Nodes and edges are typed (e.g., fork, read).
- Graph evolves from a steam of typed edges.
- Bounded space and time complexity

• Manzoor, Emaad, et al. “Fast Memory-efficient Anomaly Detection in Streaming Heterogeneous Graphs.” KDD 2016.

Graph Representation

❑ Graph to vectors via shingling

- **Key idea:** encode the heterogeneous graph into a low-dimensional representation.
- Compute the shingle vector for each graph $G^{(t)}$.
- Contain the frequencies of each k-shingle in $G^{(t)}$.



	z_G	$z_{G'}$
A	0	2
C	1	1
D	1	0
E	1	0
AxByC	1	1
BrArA	0	1
BpEoD	1	0

Figure adopted from [Manzoor et al., KDD 2016]

- Manzoor, Emaad, et al. "Fast Memory-efficient Anomaly Detection in Streaming Heterogeneous Graphs." KDD 2016.

Identification of Rare Categories

- **Sketching graphs**
  - Shingle universe is large and unknown
  - Compute *L*-dimension projection vector from shingle vector via Locality-Sensitive Hashing (LSH)

- **Detection**
  - Bootstrap K Clusters
  - Cluster Centroid: “Average” graph
  - Update Clusters: Constant time
  - Anomaly Score: Nearest centroid

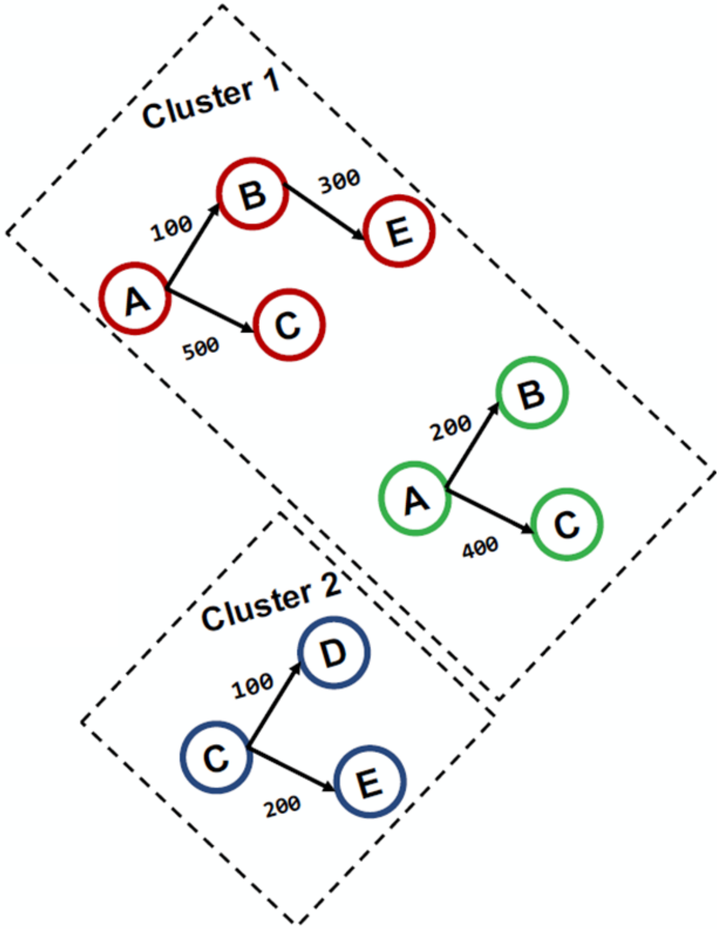


Figure adopted from [Manzoor et al., KDD 2016]

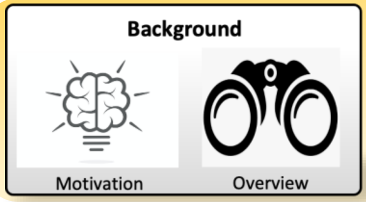
- Manzoor, Emaad, et al. “Fast Memory-efficient Anomaly Detection in Streaming Heterogeneous Graphs.” KDD 2016.

Additional References

- **Clustering-based approaches**
  - Nibble [Spielman et al., 2013]
  - TSC [Benson et al., 2015]
  - TECTONIC [Tsourakakis et al., 2017]
  - MOTLOC [Zhou et al., 2018]
  - TGS [Garranza et al., 2020]
  - ...

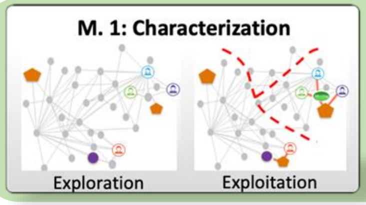
- **Graph neural network-based approaches**
  - SPNN [Meng et al., 2018]
  - hpGAT [Liu et al., 2019]
  - FI-GNNs [Ding et al., 2019]
  - Dominant [Ding et al., 2019]
  - DevNet2 [Pang et al., 2019]
  - ...

Roadmap



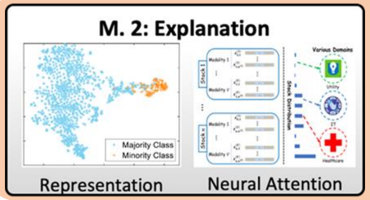
Background

- ➢ Motivation
- ➢ Research overview



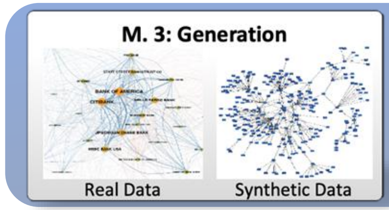
I. Rare Category Characterization

- ➢ Rare category characterization on homogeneous graphs
- ➢ Rare category characterization on heterogeneous graphs



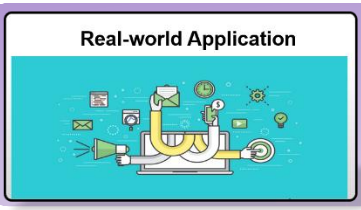
II. Rare Category Explanation

- ➢ Data insights
- ➢ Model insights



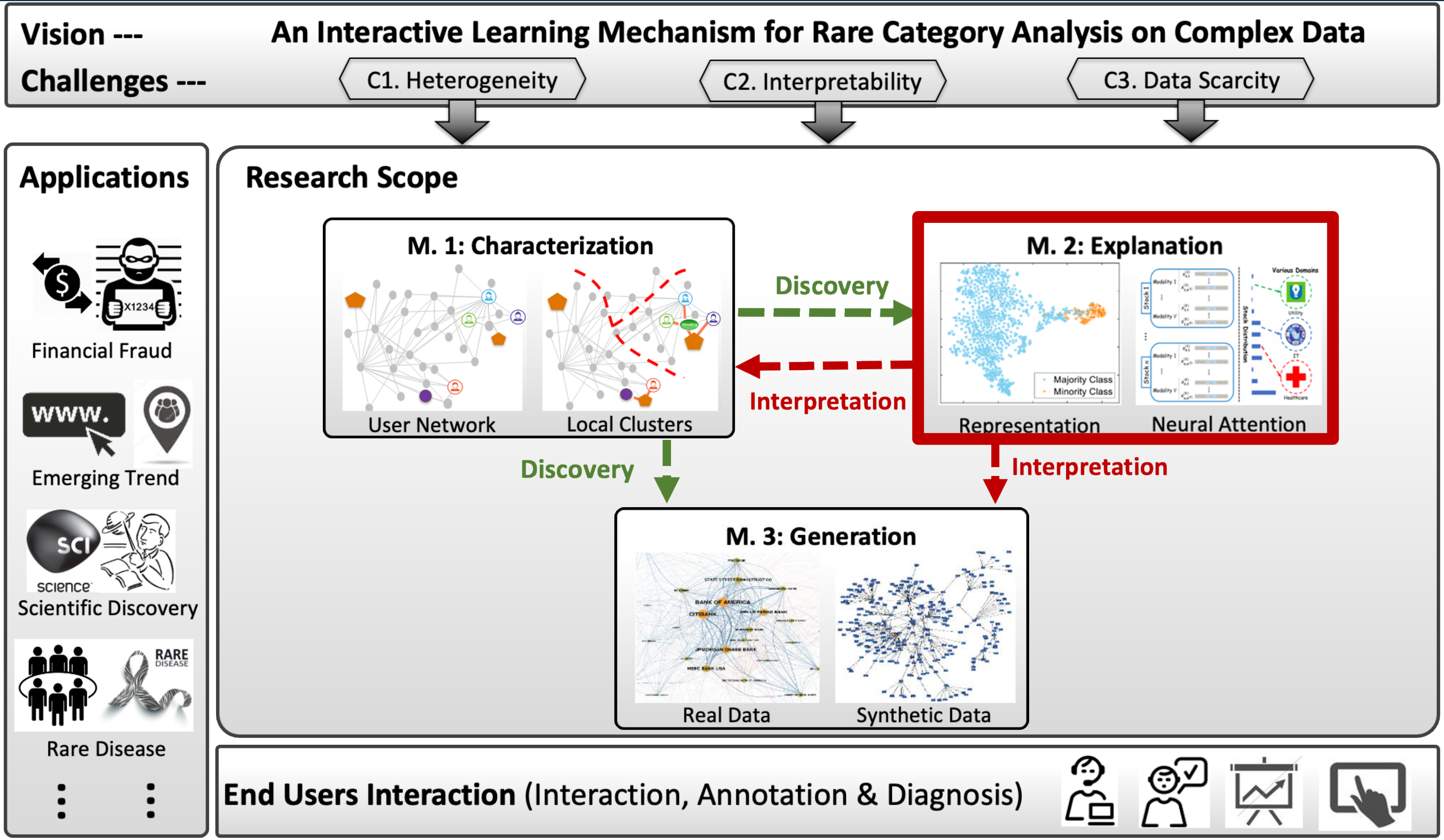
III. Rare Category Generation

- ➢ Unsupervised rare category generation
- ➢ Supervised rare category generation



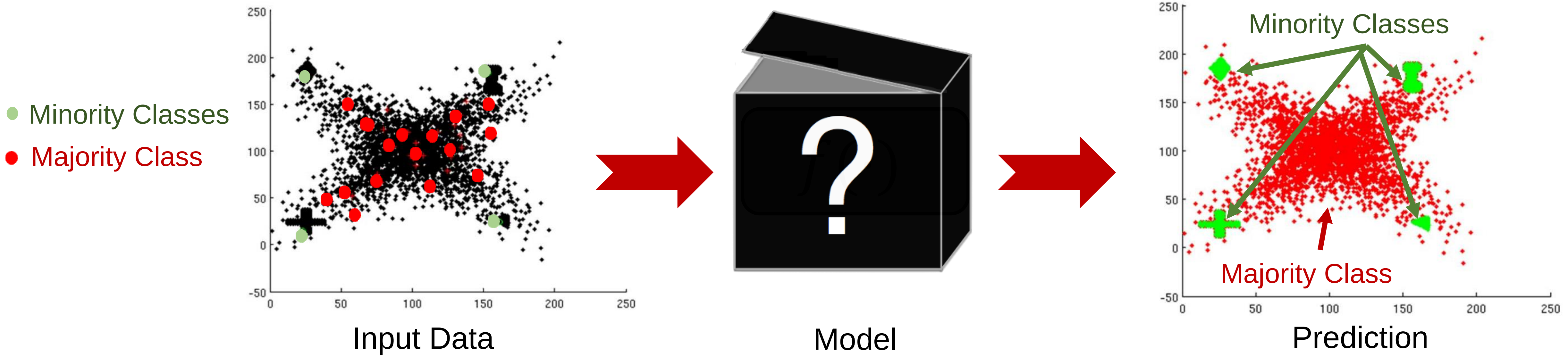
IV. Real-world Application

Rare Category Explanation



Rare Category Explanation

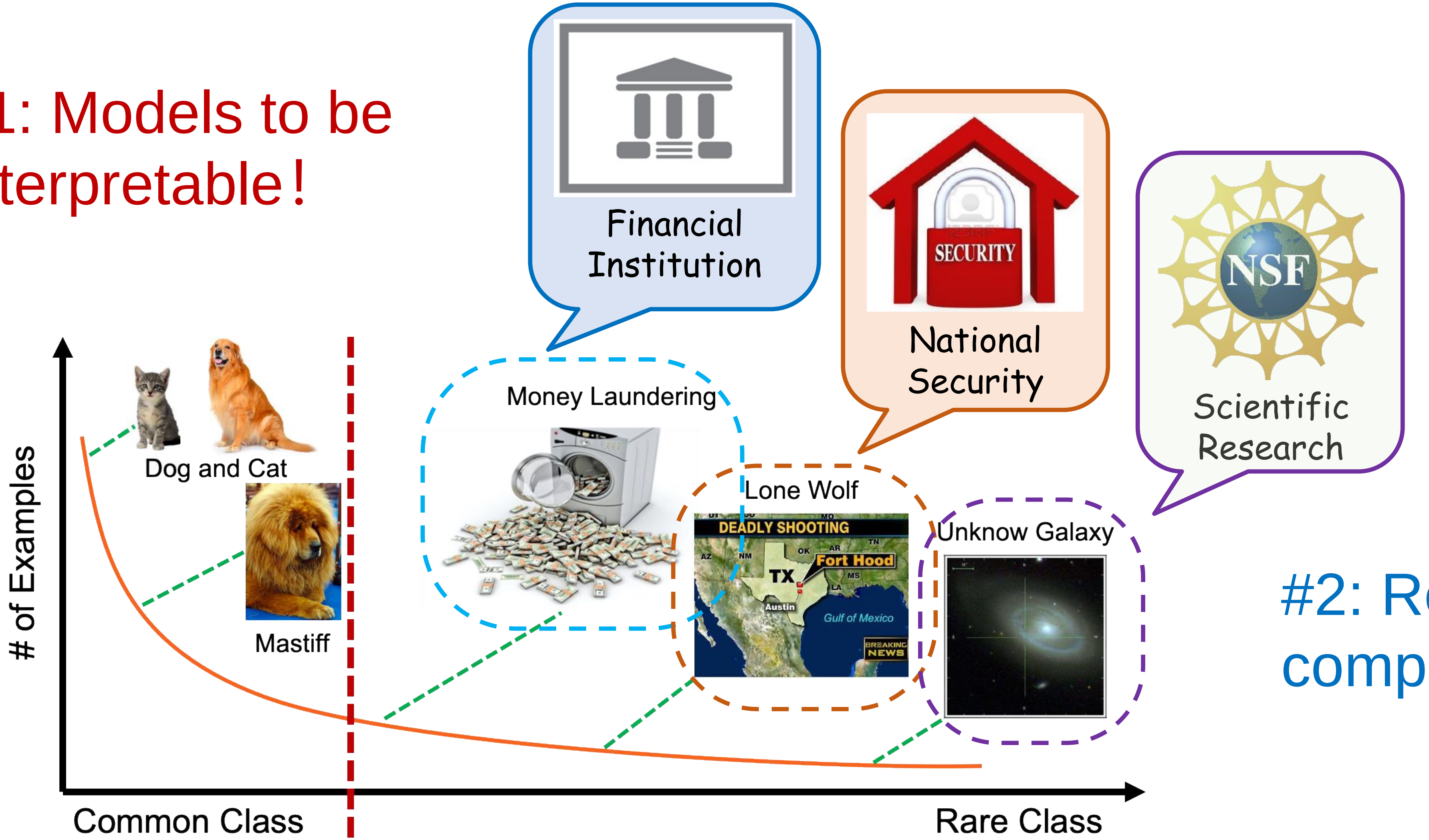
- **Problem 2.1. Rare category explanation**
  - **Given**: (1) A data set $D = \{x_1, \ldots, x_n\}$, (2) a prediction model $f(x)$ for rare category analysis in D.
  - **Find**: Interpretation over the observed data D and the prediction model $f(x)$.



- Pelleg, Dan, and Andrew W. Moore. "Active learning for anomaly and rare-category detection." Advances in neural information processing systems. 2005.
- Zhou, Dawei, et al. "A local algorithm for structure-preserving graph cut." ACM SIGKDD. 2017.

Rare Category Explanation

#1: Models to be interpretable!

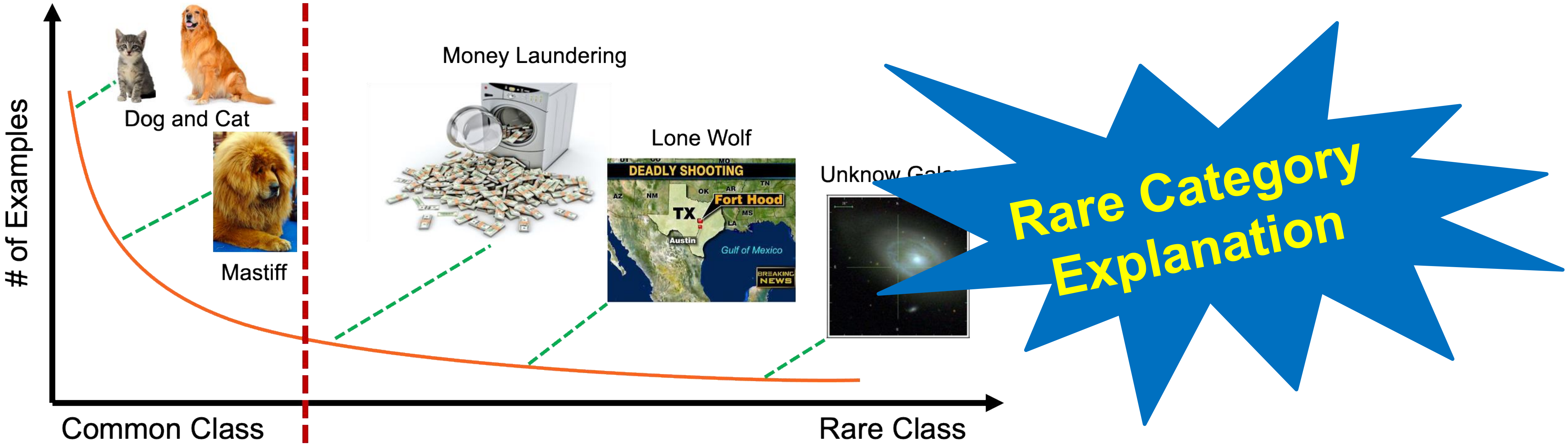


#2: Results to meet compliance!

- Abazajian, Kevork N., et al. "The seventh data release of the Sloan Digital Sky Survey." The Astrophysical Journal Supplement Series. 2009.
- Spaaij, R.. "The enigma of lone wolf terrorism: An assessment." Studies in Conflict & Terrorism. 2010.
- Fich, Eliezer M., and Anil Shivdasani. "Financial fraud, director reputation, and shareholder wealth." Journal of financial Economics. 2007.

Research Questions

- **Research questions**
  - **Q 2.1. Data insights**
    - How is the data distributed?
  - **Q 2.2. Model insights**
    - Why does the model make a certain prediction on a particular piece of information?



Rare Category Explanation

Part II: Rare Category Explanation

Data Insights

- Learning representation with imbalanced data distribution
- Learning representation with label scarcity

Model Insights

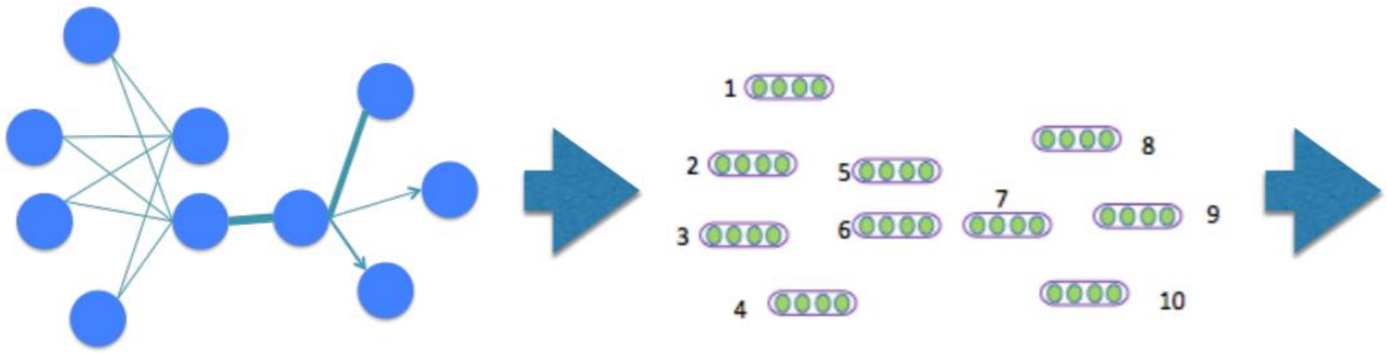
- Rare category interpretation
- Visual analytic system for rare category analysis

- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.

Network Representation Learning

❑ Learning node representation for networks

- **Goal**: Represent each node with a vector representation that preserves the structure of networks.
- **Methods**: DeepWalk [Perozzi et al., 2014], Node2vec [Grover et al., 2016], LINE [Tang et al., 2015], TransE [Bordes et al., 2013], GCN [Kipf et al., 2017], GAT [Veličković et al., 2017], GraphSAGE [Hamilton et al, 2016], etc.
- Unable to handle highly-skewed data distribution.



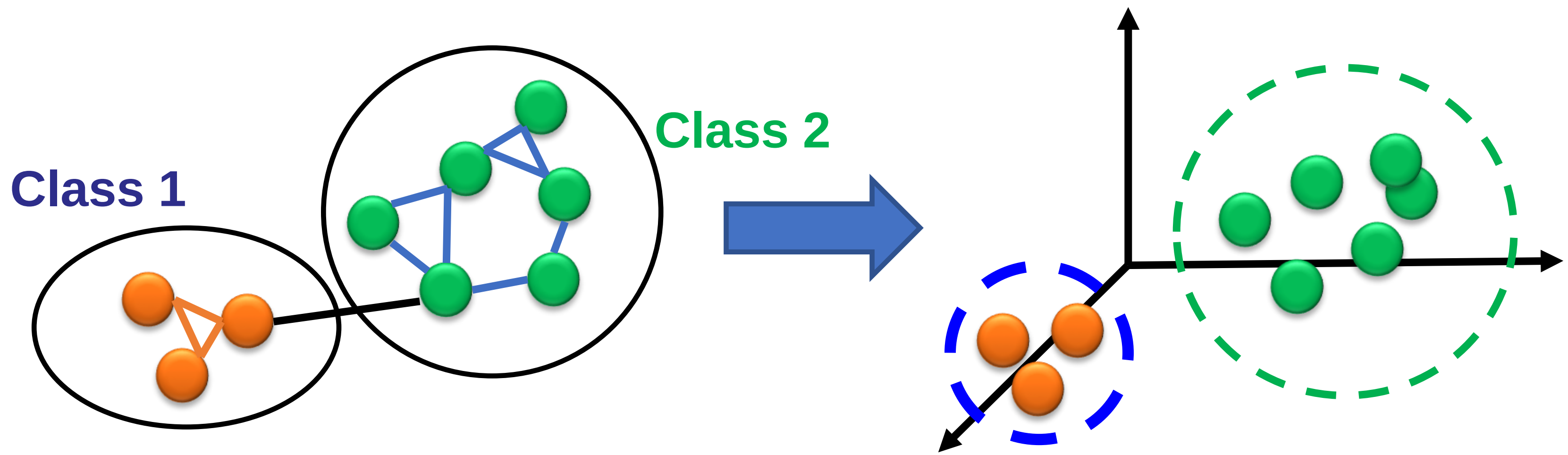
- Node classification
- Node clustering
- Link prediction
- Recommendation
- …

Network Embedding Downstream Tasks

Network Representation for Imbalanced Data

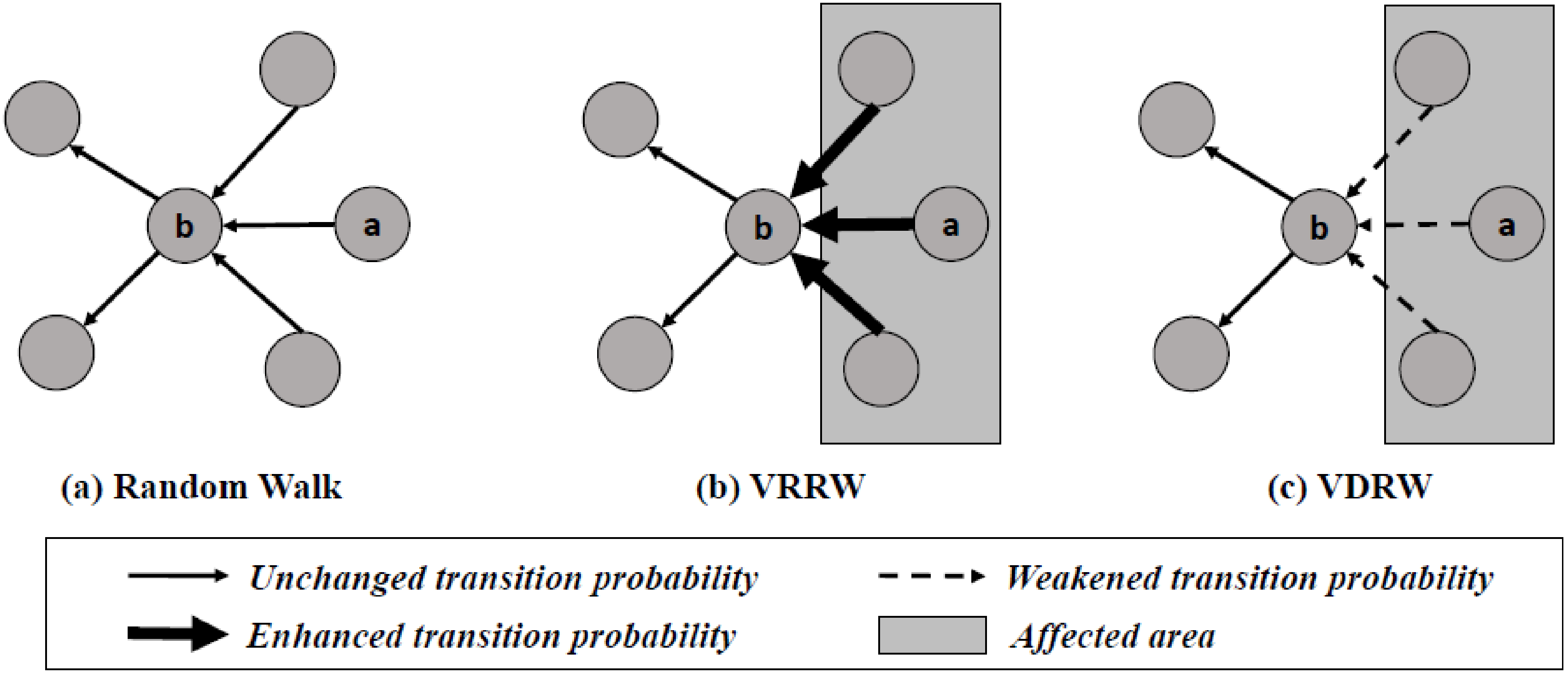
❑ Problem 2.2. Imbalanced network representation

- **Given**: (1) a (directed or undirected) network $G=(V,E)$, (2) imbalanced class labels for nodes in V, (3) embedding dimension d.
- **Find**: 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.



Vertex-Diminished Random Walk

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



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

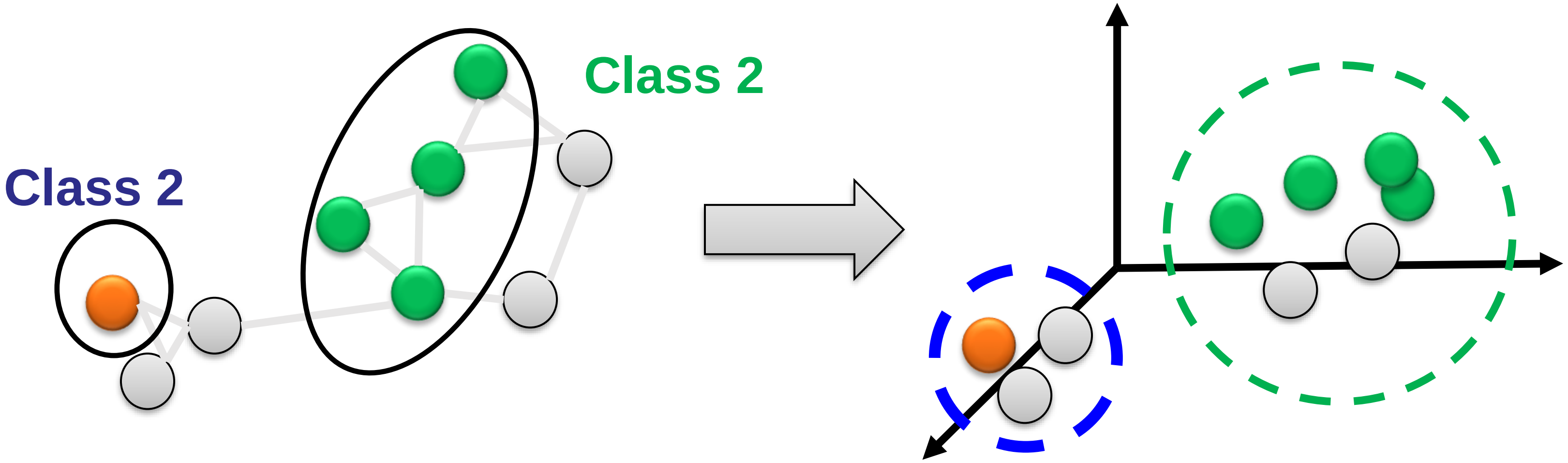
- Wu, Jun, et al. ImVerde: Vertex-Diminished Random Walk for Learning Imbalanced Network Representation. IEEE Big Data, 2018.
- Robin Pemantle. "Vertex-reinforced random walk." Probability Theory and Related Fields, 1992.

Imbalanced Network Embedding

The loss of the proposed 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



- Wu, Jun, et al. ImVerde: Vertex-Diminished Random Walk for Learning Imbalanced Network Representation. IEEE Big Data, 2018.

Imbalanced Network Embedding

❑ Unsupervised loss

Positive pair

Negative pair

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

(v_i, v_c):	a node-context pair of node i and its context c sampled from path p using the proposed VDRW
$\boldsymbol{e}_i$:	embedding vector of node i
$\boldsymbol{w}_c$:	representation of context c
v_n:	a randomly selected negative sample

❑ Supervised loss (cross-entropy error)

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

Label information

l_1, l_2:	number of hidden layers in neural networks
y_i:	class label
$\boldsymbol{x}_i$:	feature vector of node i

- Wu, Jun, et al. ImVerde: Vertex-Diminished Random Walk for Learning Imbalanced Network Representation. IEEE Big Data, 2018.

Rare Category Explanation

Part II: Rare Category Explanation

Data Insights

- Learning representation with imbalanced data distribution
- Learning representation with label scarcity

Model Insights

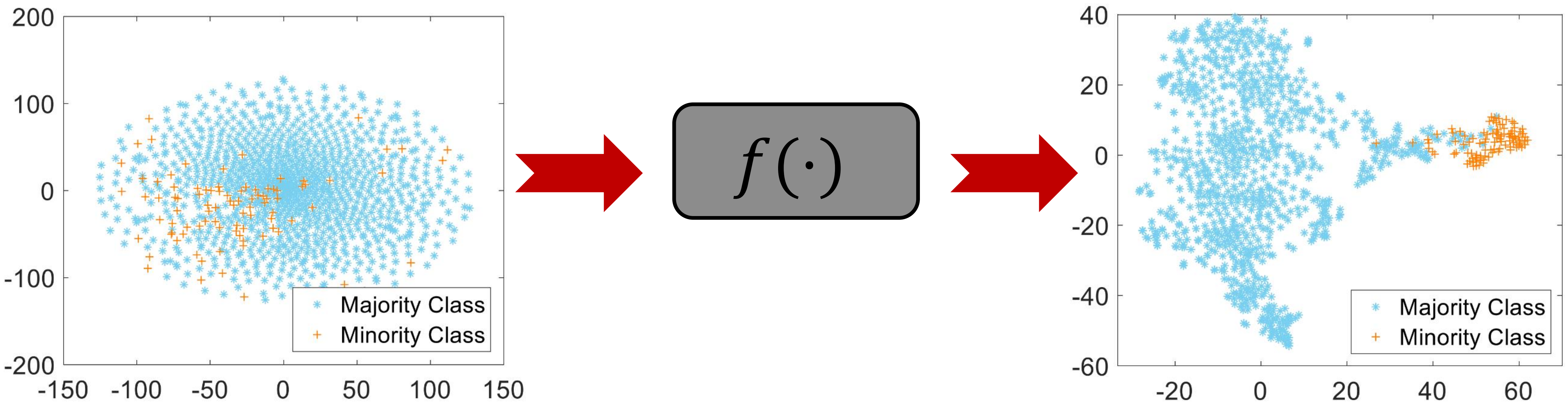
- Rare category interpretation
- Visual analytic system for rare category analysis

- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.

Learning Representation with Label Scarcity

❑ Problem 2.3. Rare-category-oriented network representation

- **Given**: (1) An attributed network $G = (V, E, \boldsymbol{X})$, (2) one-shot or few-shot labeled rare examples $L = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_L\}$, and (3) the desired embedding dimension d.
- **Find**: (1) A rare-category-oriented network embedding $\boldsymbol{E} \in R^{n \times d}$, and (2) a list of predicted rare category examples.



(a) Original Feature Space (b) Embedding Space

- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.

SPARC – Algorithm

❑ The overall optimization problem

- $\boldsymbol{L}$ denotes the set of labeled examples; $\boldsymbol{U}$ denotes the set of unlabeled examples.
- $\boldsymbol{X} = \{\boldsymbol{x_1}, \dots, \boldsymbol{x_n}\}$ denotes node attributes.
- $\boldsymbol{E} = \{\boldsymbol{e_1}, \dots, \boldsymbol{e_n}\}$ denotes the learned node embedding.
- $\boldsymbol{\theta}$ denotes the neural network parameters, and C denotes the number of classes.

$\mathcal{L}_{co}$: Self-paced learning

$$\arg\min_{\boldsymbol{E},\boldsymbol{v},\boldsymbol{\theta}}: \quad \mathcal{L} = \sum_{i=1}^{L}\sum_{c=1}^{C} c_{y_i,\hat{y}_i} \log \Pr(\hat{y}_i = c|\boldsymbol{x_i}, \boldsymbol{e_i}) - \sum_{i=1}^{L+U}\left[\sum_{c=1}^{C} \lambda^{(c)} v_i^{(c)} + \lambda^{(0)} v_i^{(0)}\right] - \alpha \sum_{i=1}^{L+U}\sum_{c=1}^{C} v_i^{(c)} v_i^{(0)}$$

$\mathcal{L}_{RCC}$: Rare category characterization $\mathcal{L}_{RCE}$: Rare category embedding

$$-\sum_{i=1}^{L+U}\sum_{c=1}^{C} v_i^{(c)} \log \Pr(\hat{y}_i = c|\boldsymbol{x_i}, \boldsymbol{e_i}) + v_i^{(0)} \log E_{(i,c,\gamma)} log\, \sigma(\gamma \boldsymbol{\theta_c^T} \boldsymbol{e_i})$$

- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.

SPARC – Algorithm

Intuitions

- Rare category characterization maximizes the probability of assigning true label to the prediction for both labeled data and unlabeled data.
- Self-paced learning starts with the easy concept and gradually learn the hard concept.
- Rare category embedding minimizes the prediction loss regarding the sampled graph context pairs.

$\mathcal{L}_{co}$: Self-paced learning

$$\arg\min_{\boldsymbol{E},\boldsymbol{v},\boldsymbol{\theta}}: \quad \mathcal{L} = \sum_{i=1}^{L}\sum_{c=1}^{C} c_{y_i,\hat{y}_i} \log \Pr(\hat{y}_i = c|\boldsymbol{x_i}, \boldsymbol{e_i}) - \sum_{i=1}^{L+U}\left[\sum_{c=1}^{C} \lambda^{(c)} v_i^{(c)} + \lambda^{(0)} v_i^{(0)}\right] - \alpha \sum_{i=1}^{L+U}\sum_{c=1}^{C} v_i^{(c)} v_i^{(0)}$$

$\mathcal{L}_{RCC}$: Rare category characterization $\quad$ $\mathcal{L}_{RCE}$: Rare category embedding

$$-\sum_{i=1}^{L+U}\sum_{c=1}^{C} v_i^{(c)} \log \Pr(\hat{y}_i = c|\boldsymbol{x_i}, \boldsymbol{e_i}) + v_i^{(0)} \log E_{(i,c,\gamma)} log\, \sigma(\gamma \boldsymbol{\theta_c^T e_i})$$

- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.

SPARC – Algorithm

❑ Rare category embedding

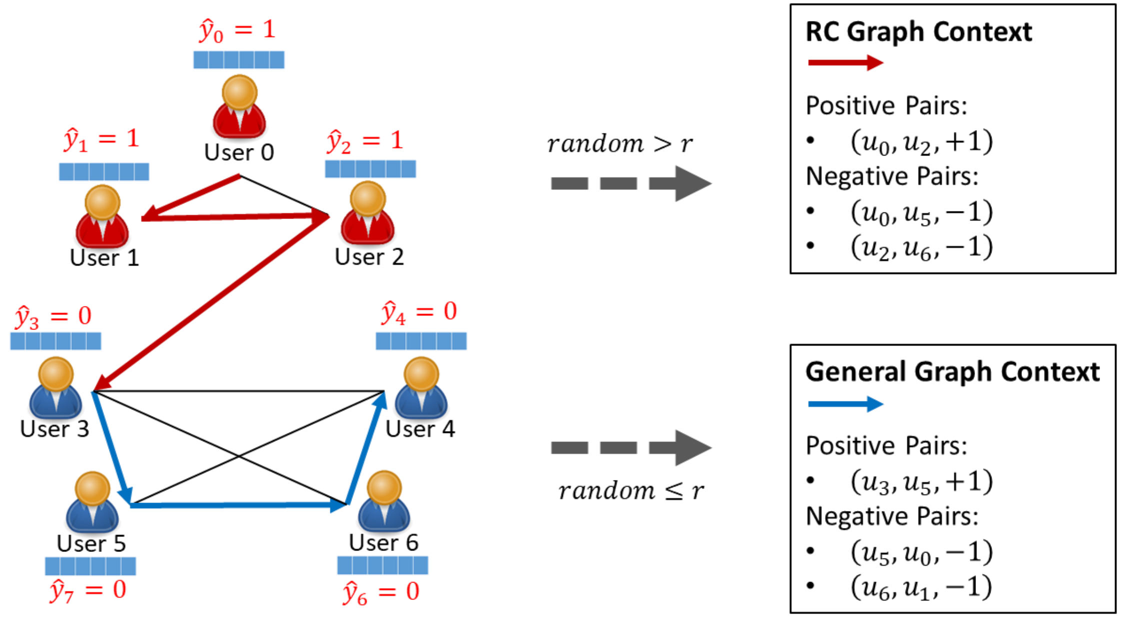
- Minimize the cross-entropy loss of predicting context pairs (i,c).

$$v_i^{(0)} \log E_{(i,c,\gamma)} log\ \sigma(\gamma \boldsymbol{\theta}_c^T \boldsymbol{e_i})$$

Indicator of positive pairs (=1) and negative pairs (=-1)

Self-paced vector for context sampling

- Self-paced dual network context sampling.
  - ➢ Extracting two types of context pairs (i,c).
  - ➢ With probability p, extracting general context pairs.
  - ➢ With probability 1−p, extracting rare-category oriented context pairs via the non-zero elements in $v^{(0)}$.



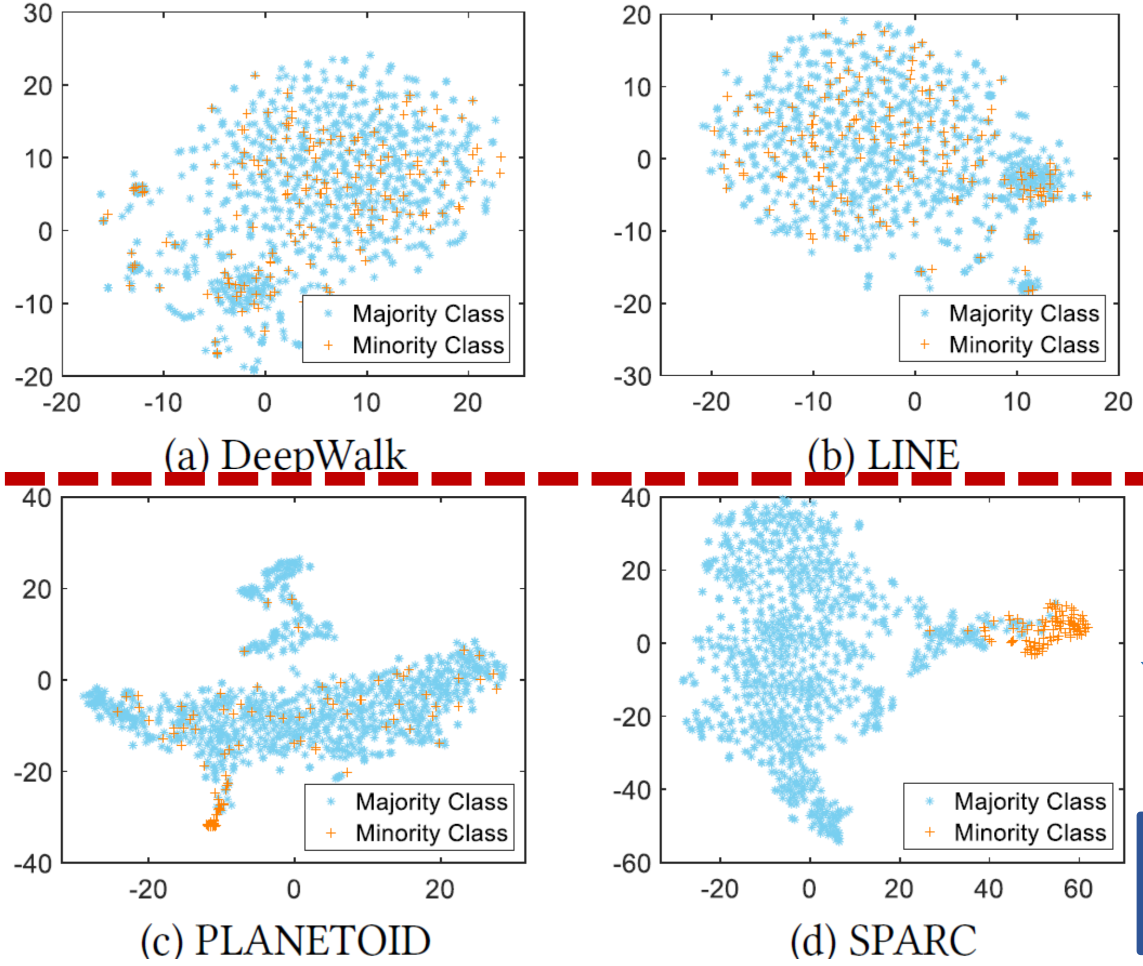
• Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.
• Kumar, M. Pawan, Benjamin Packer, and Daphne Koller. "Self-paced learning for latent variable models." Advances in neural information processing systems. 2010.

Experimental Results

❑ Network layout visualization

Unsupervised network embedding

Label-informed network embedding



Rare category is not well characterized.

Rare category is well separated from the background!

- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.
- Yang, Zhilin, et al. "Revisiting semi-supervised learning with graph embeddings." ICML. 2016.

Rare Category Explanation

Part II: Rare Category Explanation

Data Insights

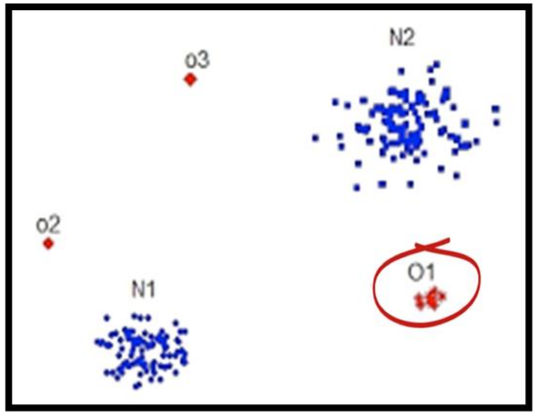
- Learning representation with imbalanced data distribution
- Learning representation with label scarcity

Model Insights

- Rare category interpretation
- Visual analytic system for rare category analysis

- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.

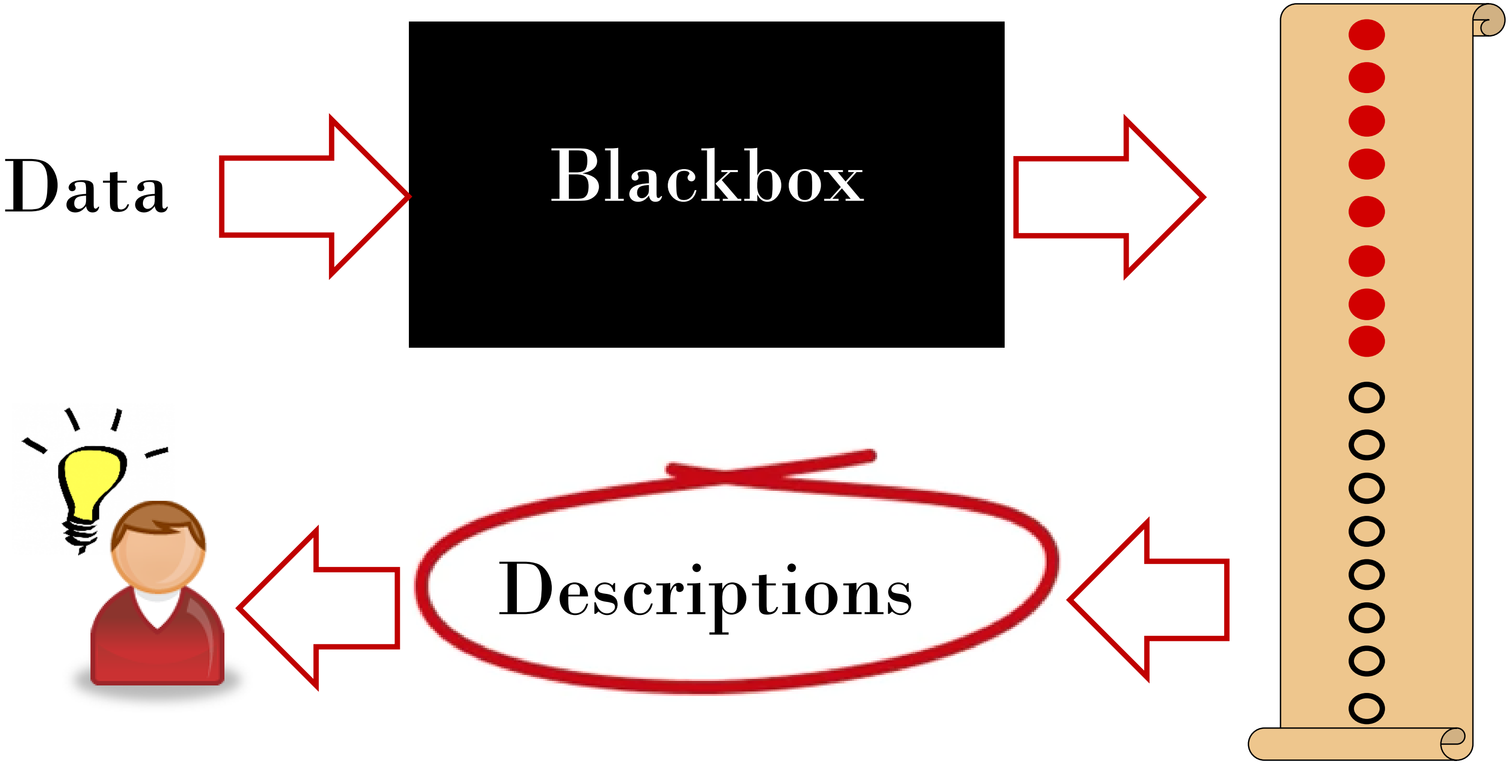
Rare Category Interpretation



Grouped Outliers = Rare Category

❑ Problem 2.4. Rare category description

- **Given**: data samples $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$.
- **Find**: The interpretation for each rare categories.



- Macha, Meghanath et al. "Explaining Anomalies in Groups with Characterizing Subspace Rules." DMKD. 2018.

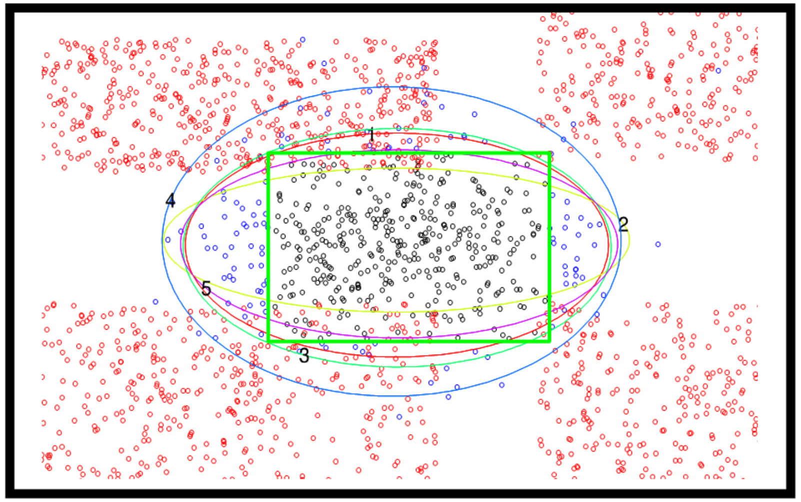
x-PACS – Algorithm

❑ Step 1: Subspace Clustering

- Hierarchically combine dense clusters.
  - Threshold mass: min # of anomalies in cluster.
- Output dense & pure subspace clusters.
  - Threshold purity: max # of normal pts in the clusters.

❑ Step 2: Rectangles to ellipsoids

- Hyper-rectangles are box-like, can refine into flexible shapes: hyper-ellipsoids.
- Recovering other nearby anomalies.



black: anomalies in subspace cluster
blue: "nearby" anomalies
red: normal pts

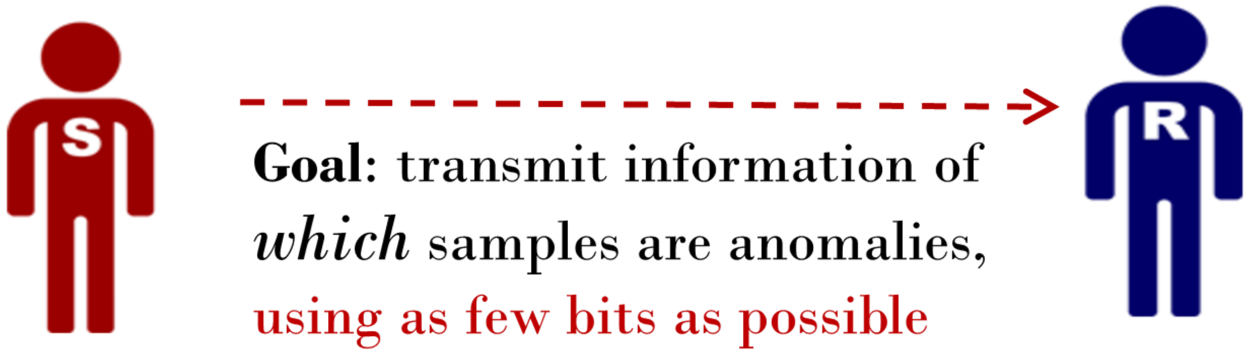
$$p(\mathbf{c},\mathbf{M}) = \{\mathbf{x} | \; (\mathbf{x}-\mathbf{c})^T \mathbf{M}^{-1} (\mathbf{x}-\mathbf{c}) \leq 1\}$$

- Macha, Meghanath et al. "Explaining Anomalies in Groups with Characterizing Subspace Rules." DMKD. 2018.

x-PACS – Algorithm

❑ Step 3: Transmission

- An Encoding Scheme:



- Generating Rules to Output

DEFINITION . (FEATURE RULES) *Given a axis-aligned ellipsoid* $p(\mathbf{c}, \mathbf{M})$ *in a subspace* $f_{t_1} \times \ldots \times f_{t_{d'}}$, *a rule on feature* t_z *is an interval* $(\mathbf{c}[z] - \text{radius}_z, \mathbf{c}[z] + \text{radius}_z)$

- Macha, Meghanath et al. "Explaining Anomalies in Groups with Characterizing Subspace Rules." DMKD. 2018.

Rare Category Explanation

Part II: Rare Category Explanation

Data Insights

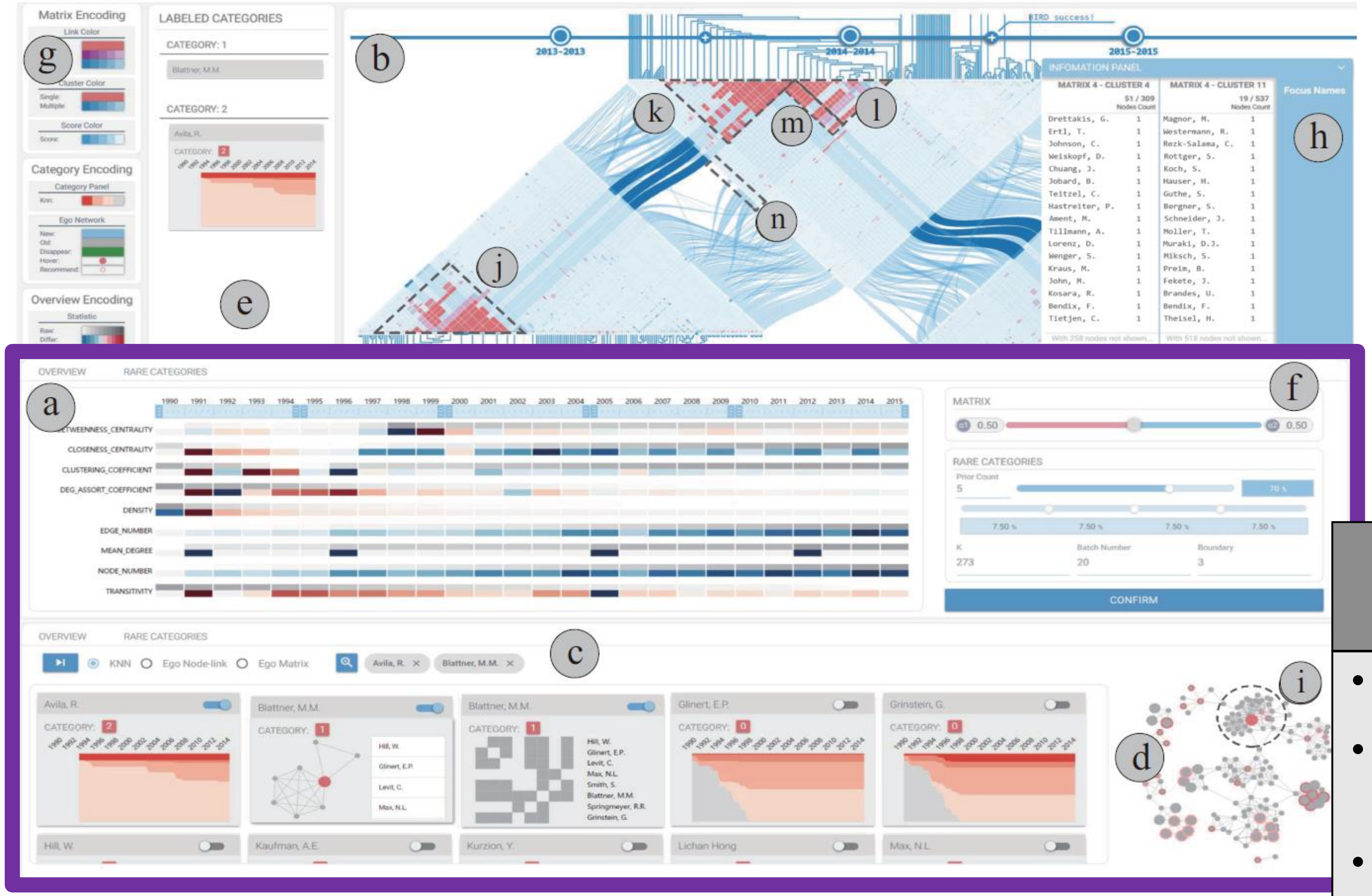
- Learning representation with imbalanced data distribution
- Learning representation with label scarcity

Model Insights

- Clustered outlier interpretation
- Visual analytic system for rare category analysis

- Zhou, Dawei, et al. "Sparc: Self-paced network representation for few-shot rare category characterization." ACM SIGKDD. 2018.

RCAnalyzer Prototype System

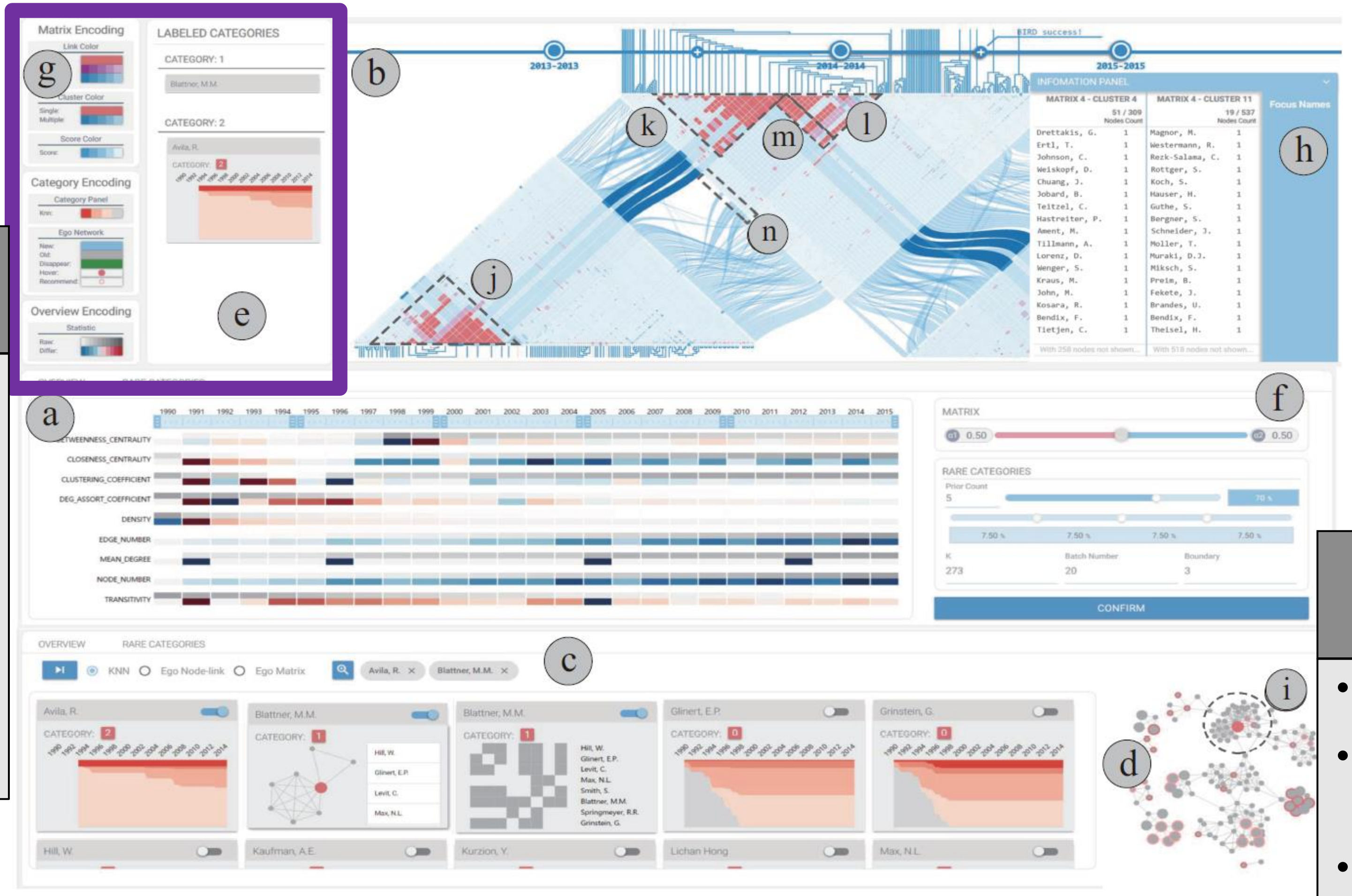


Data Exploration Module

- Represent raw data
- Interactive visualization for data querying
- Support data filtering

- Pan, Jia-cheng, et al. "RCAnalyzer: visual analytics of rare categories in dynamic networks." Frontiers of Information Technology & Electronic Engineering. 2020.
- Zhou, Dawei, et al. "Discovering rare categories from graph streams." Data mining and knowledge discovery. 2017.

RCAnalyzer Prototype System



Feature Selection Module

- Visualize the variance of data
- Visualize the correlation of data
- Guide the feature selection and subspace investigation process

Data Exploration Module

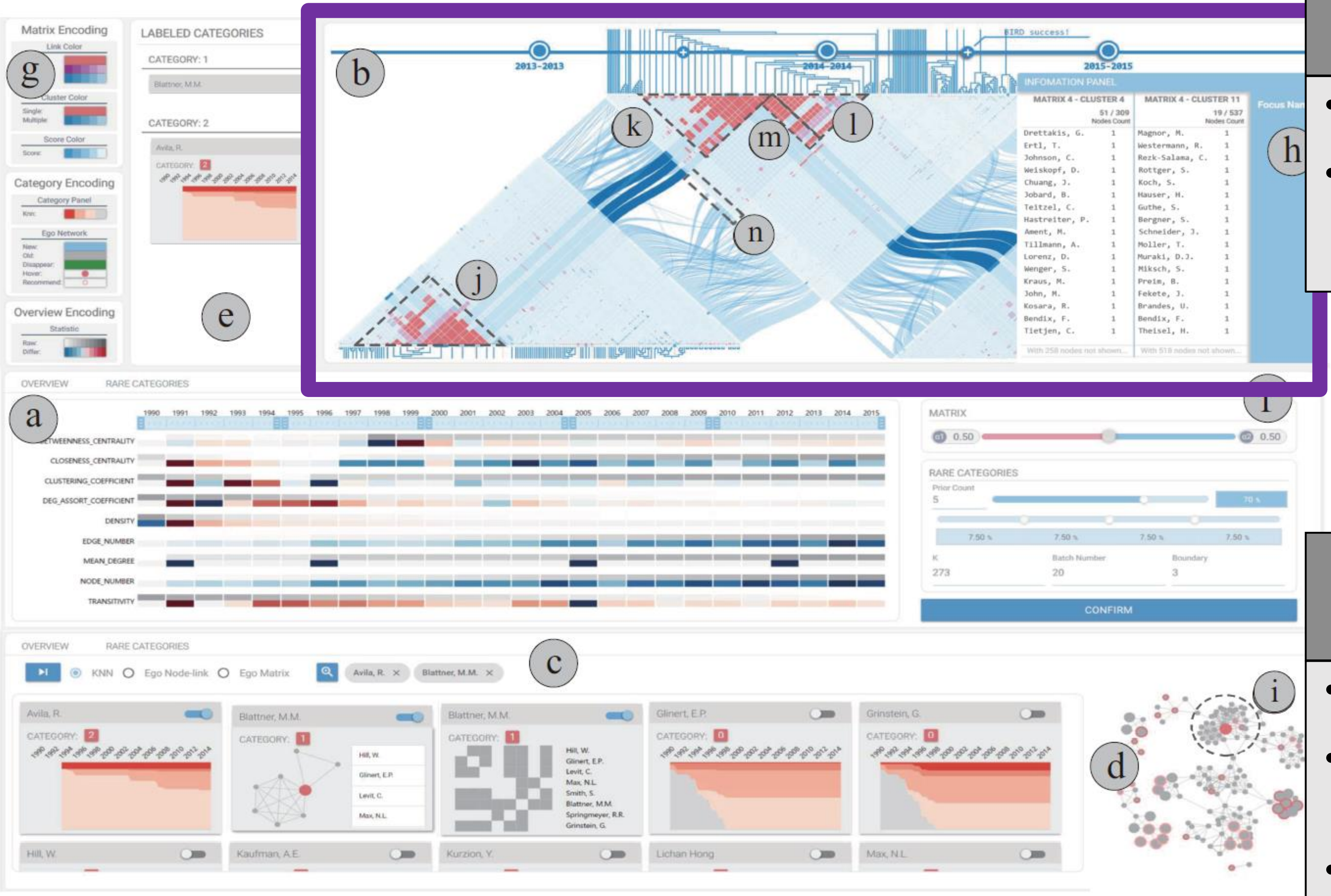
- Represent raw data
- Interactive visualization for data querying
- Support data filtering

- Pan, Jia-cheng, et al. "RCAnalyzer: visual analytics of rare categories in dynamic networks." Frontiers of Information Technology & Electronic Engineering. 2020.
- Zhou, Dawei, et al. "Discovering rare categories from graph streams." Data mining and knowledge discovery. 2017.

RCANALYZER Prototype System

Feature Selection Module

- Visualize the variance of data
- Visualize the correlation of data
- Guide the feature selection and subspace investigation process



Rare Category Analysis Module

- Interactive active learning
- Visualize rare examples in a salient representation

Data Exploration Module

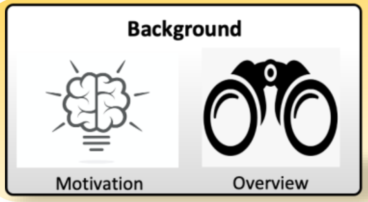
- Represent raw data
- Interactive visualization for data querying
- Support data filtering

- Pan, Jia-cheng, et al. "RCAnalyzer: visual analytics of rare categories in dynamic networks." Frontiers of Information Technology & Electronic Engineering. 2020.
- Zhou, Dawei, et al. "Discovering rare categories from graph streams." Data mining and knowledge discovery. 2017.

Additional References

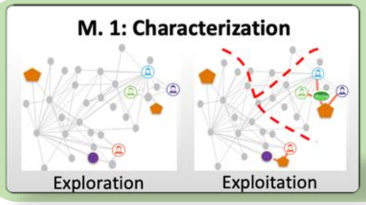
- **Representation-based approaches**
  - LMLE-kNN [Huang et al., 2016]
  - PLANETOID [Yang et al., 2016]
  - XGBOD [Zhao et al., 2018]
  - s2sL [Dumpala et al., 2018]
  - CRT [Chen et al., 2020]
  - …
- **Interpretation-based approaches**
  - COIN [Liu et al., 2018]
  - SAIFE [Rajendran et al., 2018]
  - LookOut [Gupta et al., 2019]
  - SIF [Li et al., 2020]
  - InFoRM [Kang et al., 2020]
  - …

Roadmap



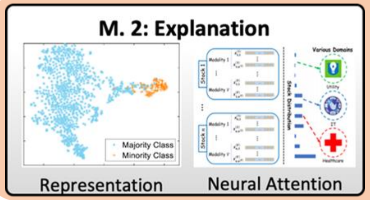
Background

- ➢ Motivation
- ➢ Research overview



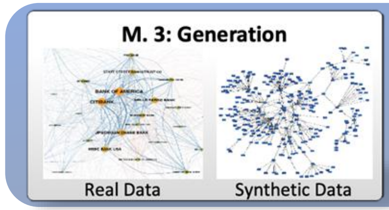
I. Rare Category Characterization

- ➢ Rare category characterization on homogeneous graphs
- ➢ Rare category characterization on heterogeneous graphs



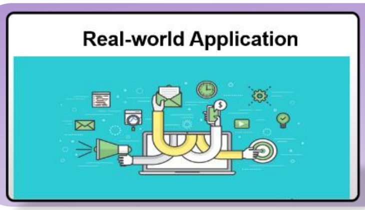
II. Rare Category Explanation

- ➢ Data insights
- ➢ Model insights



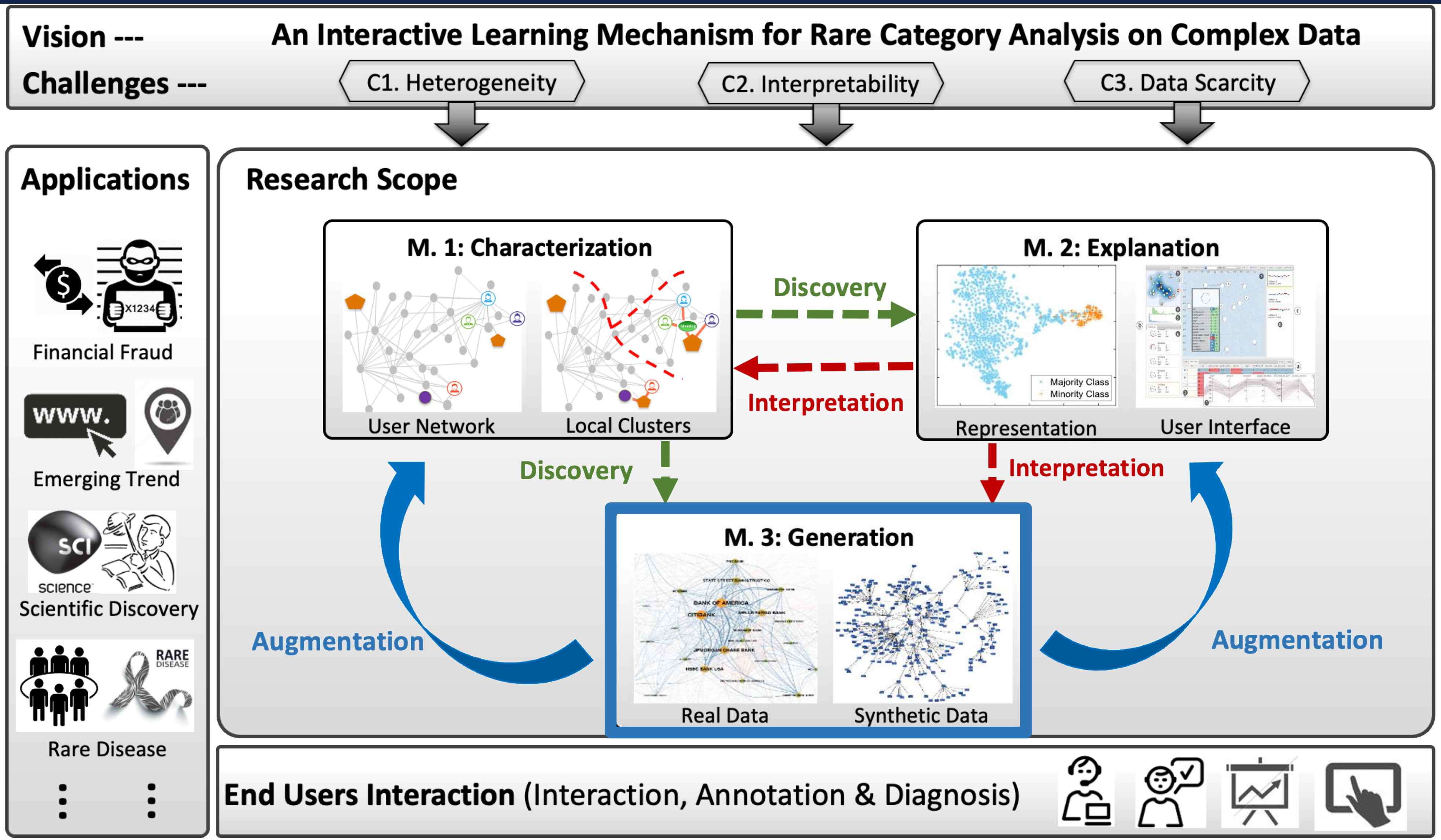
III. Rare Category Generation

- ➢ Unsupervised rare category generation
- ➢ Supervised rare category generation



IV. Real-world Application

Rare Category Generation



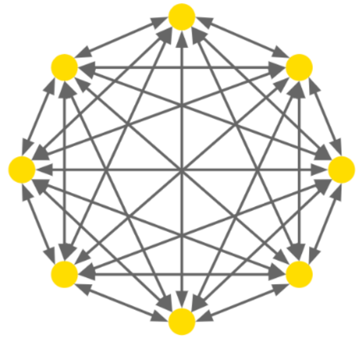
Rare Category Generation

❑ Problem 3.1. Rare category generation

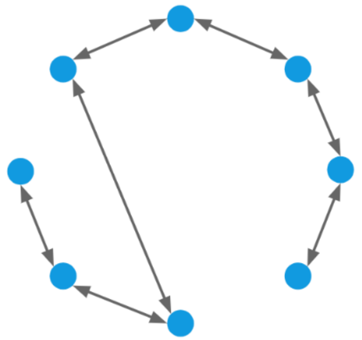
- **Given**: (1) A data set $D = \{x_1, \dots, x_n\}$, (2) a rare category analysis model $f(x)$ trained in D.
- **Find**: (1) A generative model $g(x)$ tailored for D, (2) an augmented data set D' that can boost the performance of $f(x)$ in downstream rare category analysis tasks.

❑ Previous work

- RCA models suffer from the data sparsity.
  - ➢ Sparse networks.
  - ➢ Rare category (local cluster) is not prominent.
- Synthetic minority oversampling technique (SMOTE).
  - ➢ Synthesizing new examples for the minority class.
  - ➢ Not for graph-structured data.



A subgraph



A subgraph
with missing links

Injecting rich connections

• Chawla, Nitesh V., et al. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research. 2002.

Rare Category Generation

Research questions

- **Q1. Unsupervised rare category generation**
  - Can we augment the sparse networks by injecting synthetic connections?
- **Q2. Supervised rare category generation**
  - Can we generate task-specific data sets while preserving key properties?

Observations

- **Unsupervised rare category generation**
  - Sparsity issue exists in graph data.
  - Augmenting the original network boosts the performance of rare category detection (Zhou et al., 2020).
- **Supervised rare category generation**
  - Label-informed Molecular graph generative model preserving key properties (Jin et al., 2018);
  - Generating molecular graphs maintaining chemical validity and high potency (Jin et al., 2018);
  - Generating molecular graphs from a specific class (Simonovsky et al., 2018);
  - Considering the desired chemical properties as the label for graph generation.

- Simonovsky, Martin et al. "Graphvae: Towards generation of small graphs using variational autoencoders." ICANN 2018.
- Jin, Wengong, et al. "Junction tree variational autoencoder for molecular graph generation." ICML 2018.
- Zhou, Dawei, et al. "A Data-Driven Graph Generative Model for Temporal Interaction Networks." ACM SIGKDD. 2020.

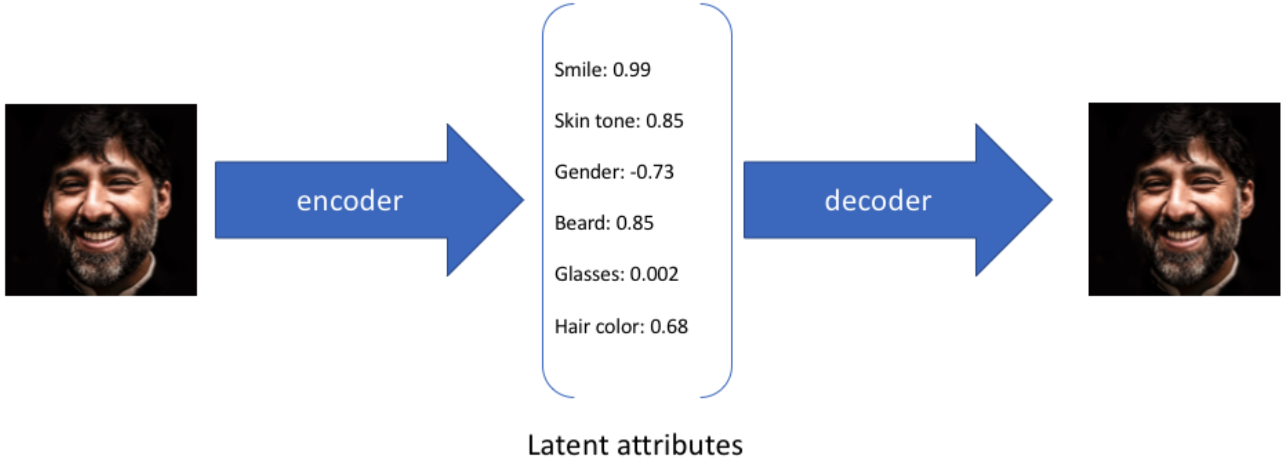
Generative Models

- **Variational autoencoder based methods**
  - GraphVAE (Simonovsky et al., 2018), JTVAE (Jin et al., 2018), Graphite-VAE (Grover et al., 2019), Reg-VAE (Ma et al., 2018), NEVAE (Samanta et al., 2019), ……
- **Generative adversarial network based methods**
  - NetGAN (Bojchevski et al., 2018), MolGAN (De Cao et al., 2018), GCPN (You et al., 2018), Misc-GAN (Zhou et al., 2019), LGGAN (Fang et al,. 2019), ……
- **Autoregressive based methods**
  - GraphRNN (You et al., 2018), TagGen (Zhou et al., 2020), MolecularRNN (Popova et al., 2019), ……

Variational Autoencoder

❑ Autoencoder

- An encoder model $q_{\phi}(z|x)$,
- A decoder model $p_{\theta}(x|z)$.



❑ Maximizing the variational lower-bound

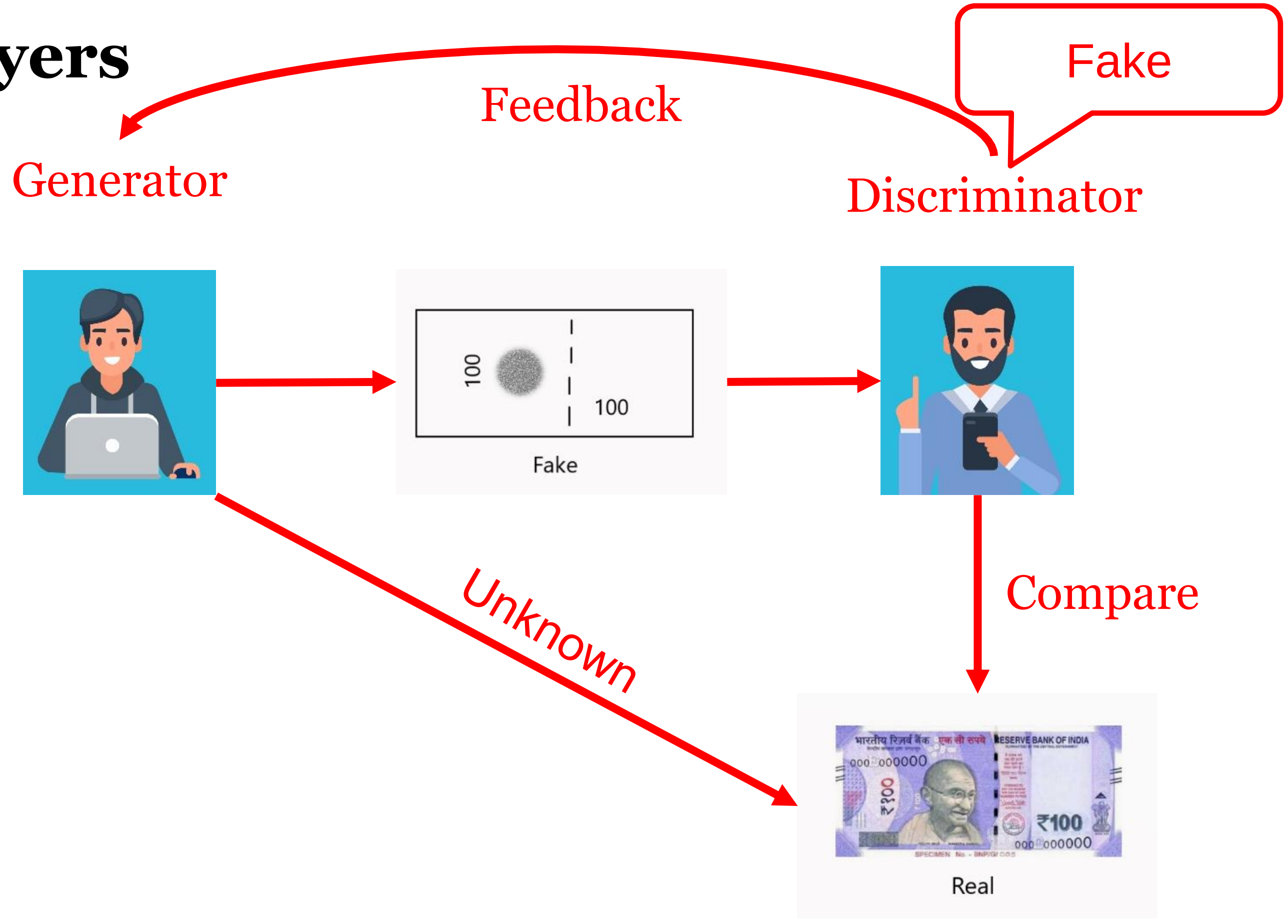
- Reparameterization trick to approximate intractable posterior

$$\mathcal{L}(\theta, \phi, x) = \underbrace{\mathbb{E}_{q_{\phi}(z|x)}\left[\log p_{\theta}(x|z)\right]}_{\text{Reconstruction Capability}} - \underbrace{D_{KL}(q_{\phi}(z|x)||p_{\theta}(z))}_{\text{Regularization}}$$

- Kingma et al. "Auto-Encoding Variational Bayes " NIPS 2014.
- Image source: https://www.jeremyjordan.me/variational-autoencoders/

Generative Adversarial Nets (GANs)

- **GAN involves two competing players**
  - Generator G
  - Discriminator D

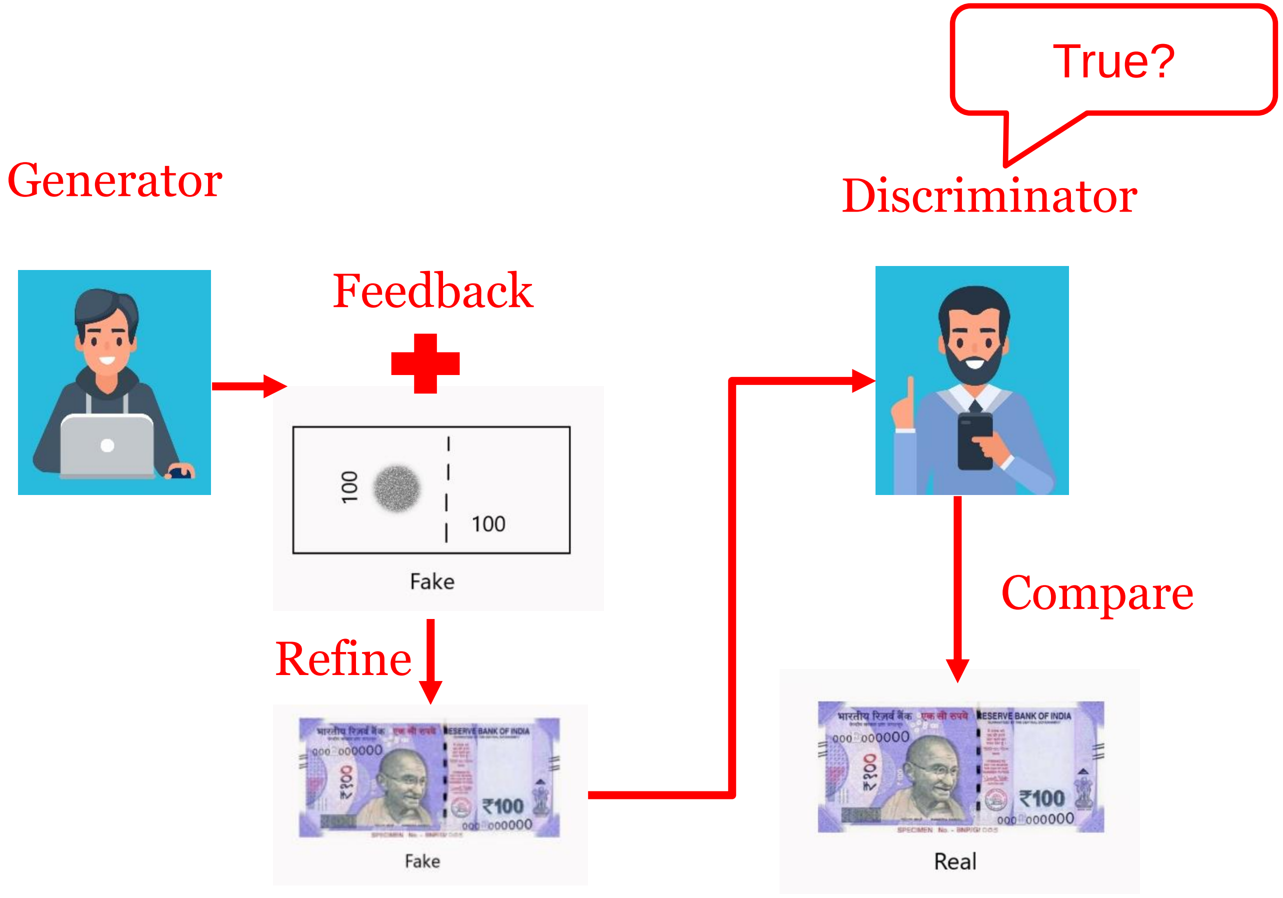


- Goodfellow, Ian, et al. "Generative adversarial nets." NIPS 2014.

Generative Adversarial Nets (GANs)

- **Generator *G***
  - Takes random variable as an input
  - Generates synthetic samples
  - Aims fool the discriminator D.

- **Discriminator *D***
  - Aims to distinguish between real data and generated data.



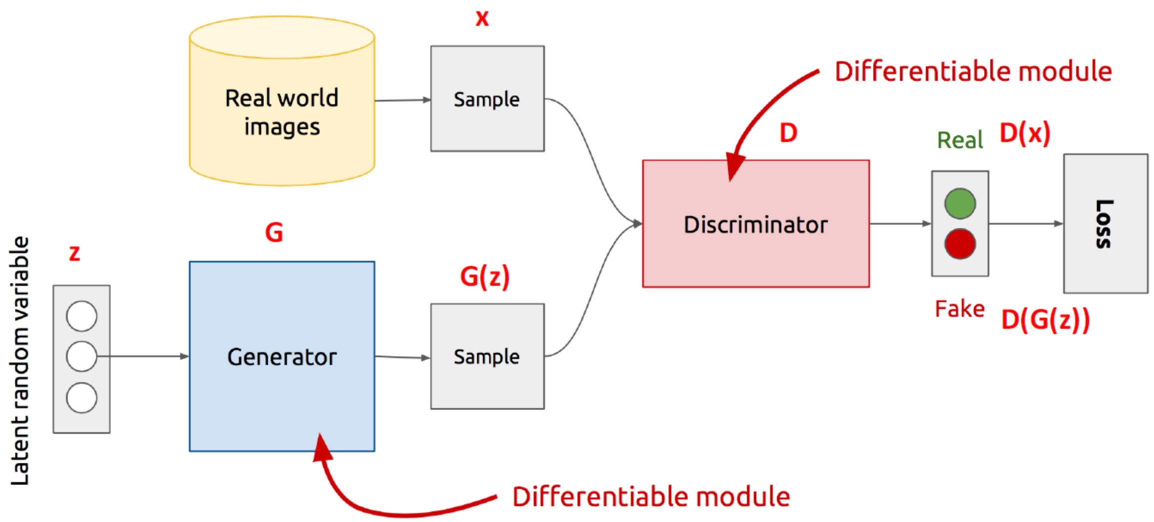
- Goodfellow, Ian, et al. "Generative adversarial nets." NIPS 2014.

Generative Adversarial Nets (GANs)

- **Generator G**
  - Takes random variable as an input
  - Generates synthetic samples
  - Aims fool the discriminator D.

- **Discriminator D**
  - Aims to distinguish between real data and generated data.

- The **min-max game** is formulated as



$$\min_G \max_D V(G, D) = \mathbb{E}_{x \sim p_{data}(x)} \log D(x) + \mathbb{E}_{z \sim p_z(z)} \log(1 - D(G(z)))$$

- Goodfellow, Ian, et al. "Generative adversarial nets." NIPS 2014.
- Image source: https://www.slideshare.net/xavigiro/deep-learning-for-computer-vision-generative-models-and-adversarial-training-upc-2016

Deep Autoregressive Models

❑ Key idea:

- Predicts future behavior based on past behavior

❑ Example

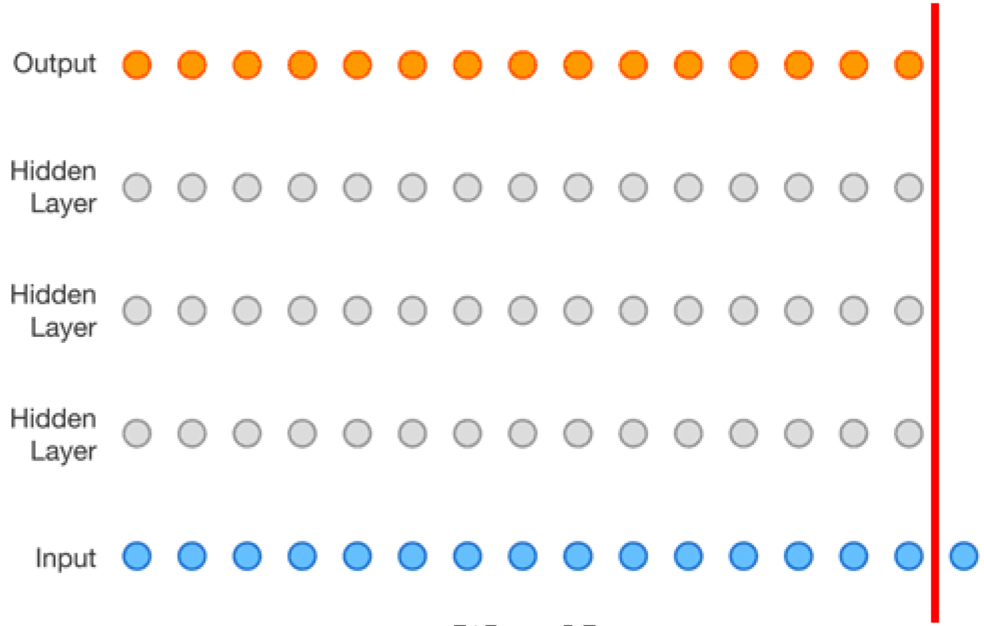
- Recurrent Neural Network (RNN)

❑ WaveNets (Oord et al. 2016)

- A generative model of audio waveforms
- $p(x) = \sum_{t=1}^{T} p(x_t | x_1, \ldots, x_{t-1})$

❑ XLNet (Yang, et al. 2019)

- A generalized autoregressive language model
- $\max_{\theta} \sum_{t=1}^{T} \log p_{\theta}(x_t | x_1, \ldots, x_{t-1})$



WaveNets

- Oord, Aaron van den, et al. "Wavenet: A generative model for raw audio." arXiv preprint arXiv:1609.03499 (2016).
- Yang, Zhilin, et al. "Xlnet: Generalized autoregressive pretraining for language understanding." NeurIPS 2019.
- Image source: https://deepmind.com/blog/article/wavenet-generative-model-raw-audio

Rare Category Generation

Part III: Rare Category Generation

Unsupervised Rare Category Generation

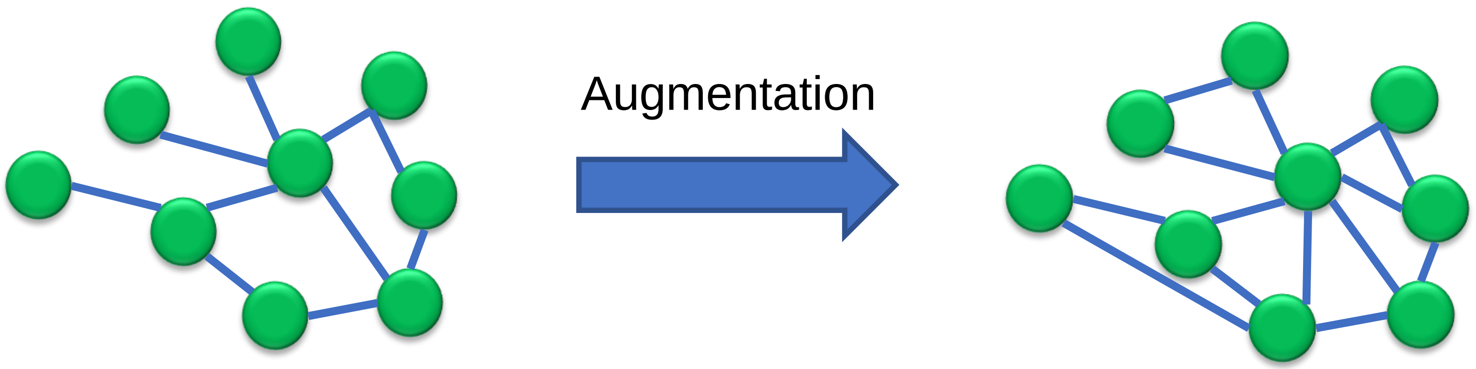
- Learning data distribution of static network without constraints
- Learning data distribution of temporal network without constraints

Supervised Rare Category Generation

- Learning data distribution of network with constraints

Unsupervised Rare Category Generation

- **Problem 3.2. Static network generation**
  - **Given**:
    - (1) A static network G,
    - (2) A rare category analysis model $f(G)$ trained in G.
  - **Find**:
    - (1) A generative model $g(G)$ tailored for G,
    - (2) An augmented graph G' that can boost the performance of $f(G')$ in downstream rare category analysis tasks.



GraphRNN

❑ Autoregressive Generative Model

- Not learns the distribution of graph $p(G)$ directly;
- Aims to learn the distribution of graph sequence $p(S^{\pi})$;

❑ Deep Autoregressive model

- One process that generates a sequence of nodes (graph-level RNN);
- One process that generates a sequence of edges for each newly added node (edge-level RNN).
- $p(S_i^{\pi}|S_{<i}^{\pi})$ captures how current node is connected to previous nodes based on how the previous nodes are connected to each other.

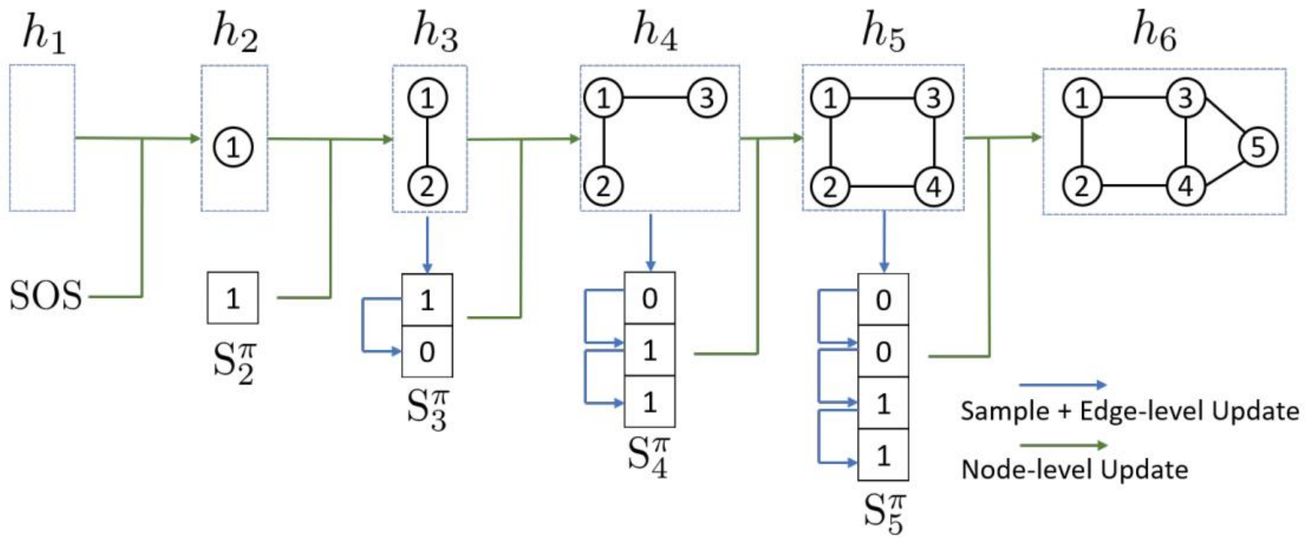
$$p(S^{\pi}) = \prod_{i=1}^{n+1} p(S_i^{\pi}|S_1^{\pi}, \ldots, S_{i-1}^{\pi}) = \prod_{i=1}^{n+1} p(S_i^{\pi}|S_{<i}^{\pi})$$

- You, Jiaxuan, et al. "Graphrnn: Generating realistic graphs with deep auto-regressive models." ICML 2018.

GraphRNN

❑ A Hierarchical RNN

- An RNN that consists of a state-transition function f_{trans} and an output function f_{out} to parameterize $p(S_i^\pi | S_{<i}^\pi)$.
- Graph-level: $h_i = f_{trans}(h_{i-1}, S_{i-1}^\pi)$ encodes the state of the graph generated so far.
- Edge-level: $f_{out}(h_i)$ generates the edges of a given node.
- Graph is assembled from $S^\pi = \{S_1^\pi, S_2^\pi, \dots, S_N^\pi\}$.



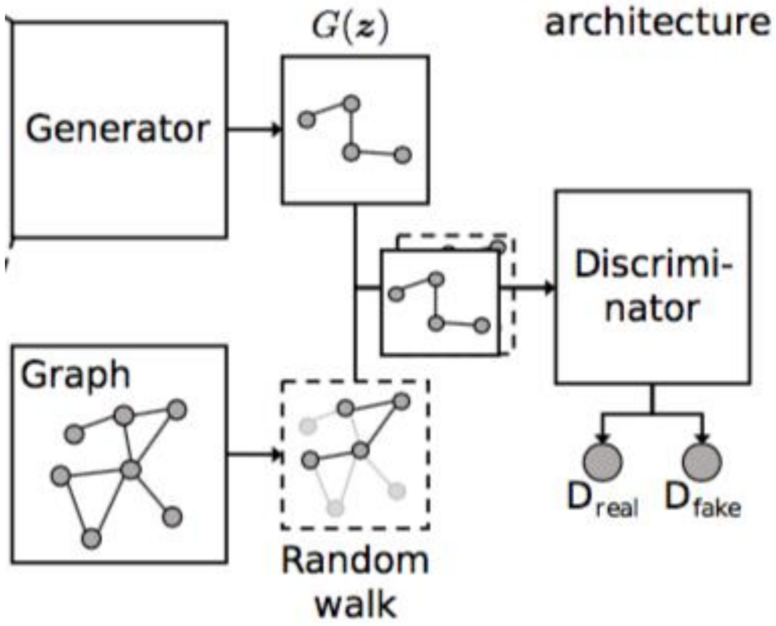
- You, Jiaxuan, et al. "Graphrnn: Generating realistic graphs with deep auto-regressive models." ICML 2018.

NetGAN

- **Key idea:**
  - Aims to learn the distribution of graph sequence

- **Step 1: Learning how to generate realistic random walks**
  - **Generator *G***
    - Generate fake random walks;
    - Learns to generate random walks $(v_1, \dots, v_T)$.
  - **Discriminator *D***
    - Takes random walks as inputs;
    - Distinguishes whether random walks are real or fake.

NetGAN architecture



- Bojchevski, Aleksandar, et al. "Netgan: Generating graphs via random walks." ICML 2018.

NetGAN

Step 2: Assembling the adjacency matrix

- Constructs a score matrix S with edge counts;
- Converts the matrix S into binary adjacency matrix A in two steps.
  - (1). Ensure that every node i has at least one edge by sampling a neighbor j with probability $p_{ij} = \frac{s_{ij}}{\sum_v s_{iv}}$.
  - (2). Continue sampling edges without replacement using the probability $p_{ij} = \frac{s_{ij}}{\sum_{u,v} s_{uv}}$ for each edge (i, j), until we reach the desired number of edges.

- Bojchevski, Aleksandar, et al. "Netgan: Generating graphs via random walks." ICML 2018.

Rare Category Generation

Part III: Rare Category Generation

Unsupervised Rare Category Generation

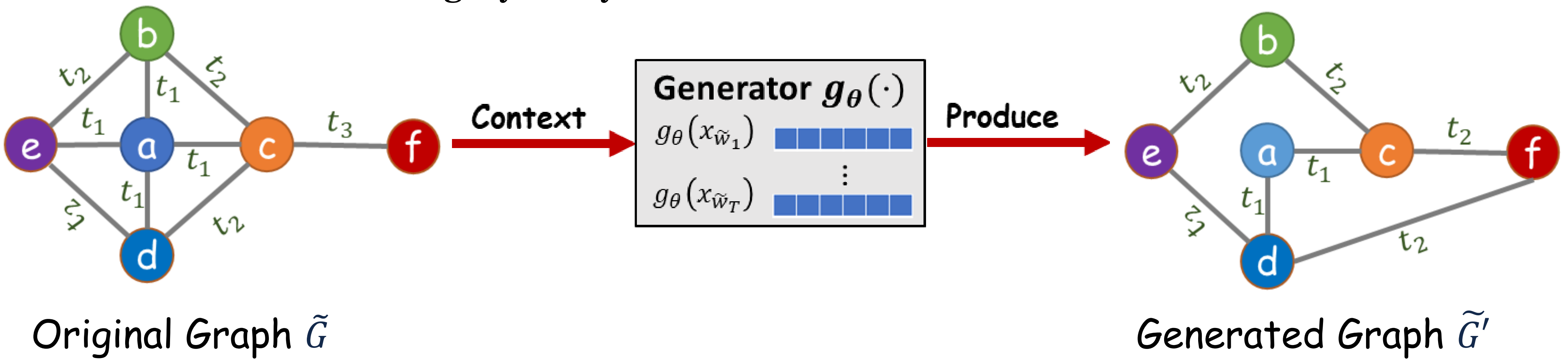
- Learning data distribution of static network without constraints
- Learning data distribution of temporal network without constraints

Supervised Rare Category Generation

- Learning data distribution of network with constraints

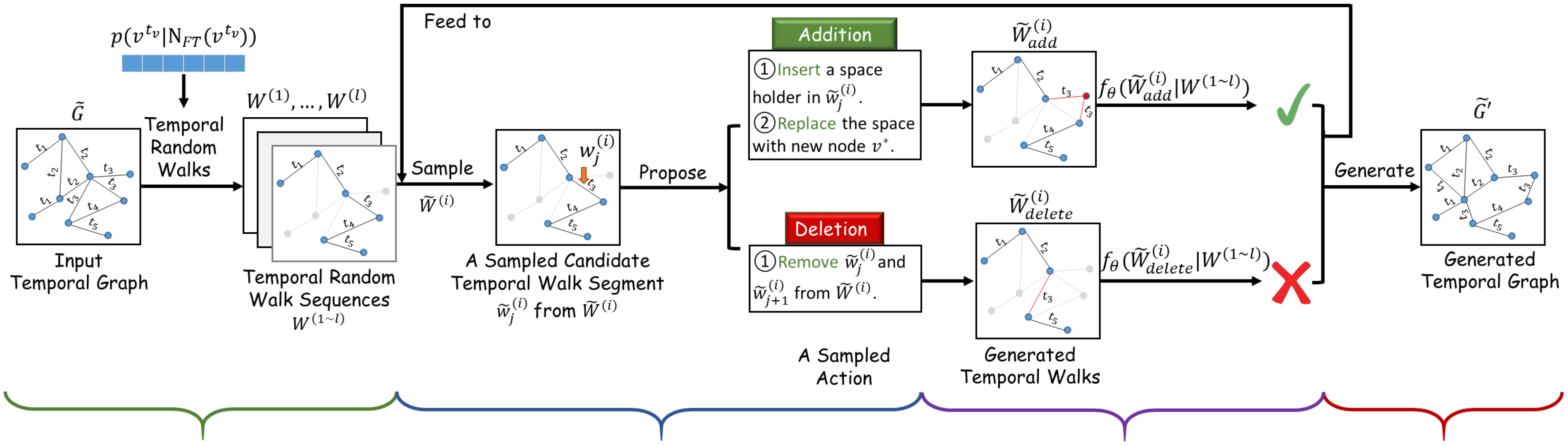
Unsupervised Rare Category Generation

- **Problem 3.3. Temporal network generation**
  - **Given:**
    - (1) A temporal network $\tilde{G}$, which is represented as a collection of timestamped edges $\tilde{E} = (e_1^{t_{e_1}}, e_2^{t_{e_2}}, \ldots, e_m^{t_{e_m}})$.
    - (2) A rare category analysis model $f(\tilde{G})$ trained in $\tilde{G}$.
  - **Find:**
    - A synthetic temporal network $\tilde{G}'$ that can boost the performance of $f(\tilde{G}')$ in downstream rare category analysis tasks.



- Zhou, Dawei, et al. "A Data-Driven Graph Generative Model for Temporal Interaction Networks." ACM SIGKDD. 2020.

TAGGEN – Temporal Graph Generation



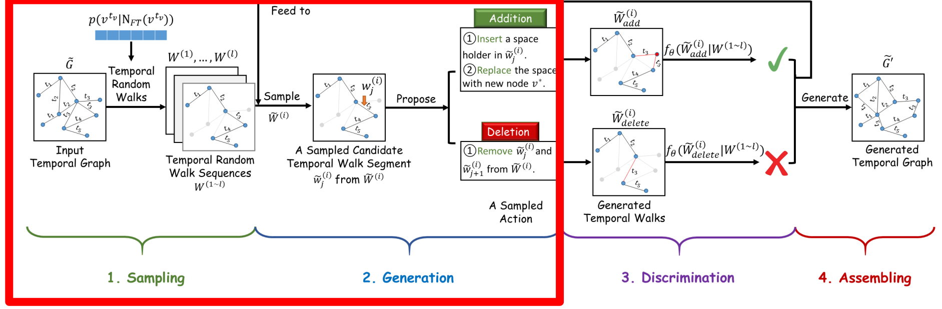
1. A novel context extraction strategy for temporal networks.

2. A family of local operations to perform addition and deletion of nodes and edges.

3. A bi-level self-attention mechanism.

- Zhou, Dawei, et al. "A Data-Driven Graph Generative Model for Temporal Interaction Networks." ACM SIGKDD. 2020.

TagGen – Algorithm



❑ S1: Context sampling

- **Goal**: selecting initial nodes for conducting temporal random walks.
- **Assumption**: weak dependence.

Temporal neighborhood of v^{t_v}

$$p(v^{t_v}|N_{FT}(v^{t_v})) > \delta[p(v^{t_v}|N_S(v^{t_v}))p(v^{t_v}|N_T(v^{t_v}))]$$

Weak dependence

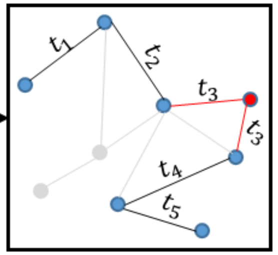
❑ S2: Sequence generation

- **Goal:** generating synthetic temporal random walks.
- **Solution:** mimicking dynamic network evolution via local operations.

Addition

①Insert a space holder in $\tilde{w}_j^{(i)}$.
②Replace the space holder with a new temporal role v^t.

$\widetilde{W}_{add}^{(i)}$

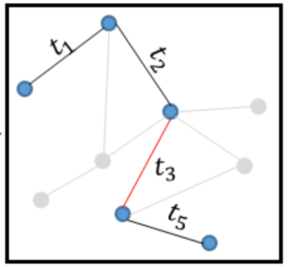


Add a temporal node

Deletion

①Remove $\tilde{w}_j^{(i)}$ and $\tilde{w}_{j+1}^{(i)}$ from $\widetilde{W}^{(i)}$.

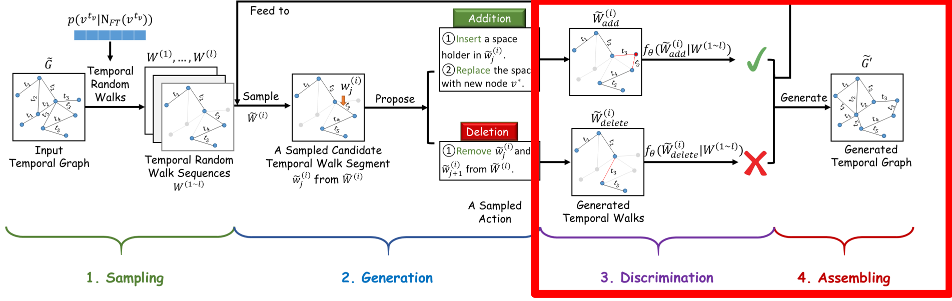
$\widetilde{W}_{delete}^{(i)}$



Remove a temporal node

- Zhou, Dawei, et al. "A Data-Driven Graph Generative Model for Temporal Interaction Networks." ACM SIGKDD. 2020.

TagGen – Algorithm



❑ S3: Sequence discrimination

- **Goal:** selecting synthetic random walks that are plausible in the input graph.
- **Solution:** a bi-level self-attention mechanism.
  - ➢ maximizing the action likelihood via the deep autoregressive model $f_\theta(\cdot)$.

$$p(\widetilde{W}^{(i)}{}_{action} | W^{(1\sim l)})$$

  - ➢ $\widetilde{W}_{action}$: generated random walk sequence after a sampled action.

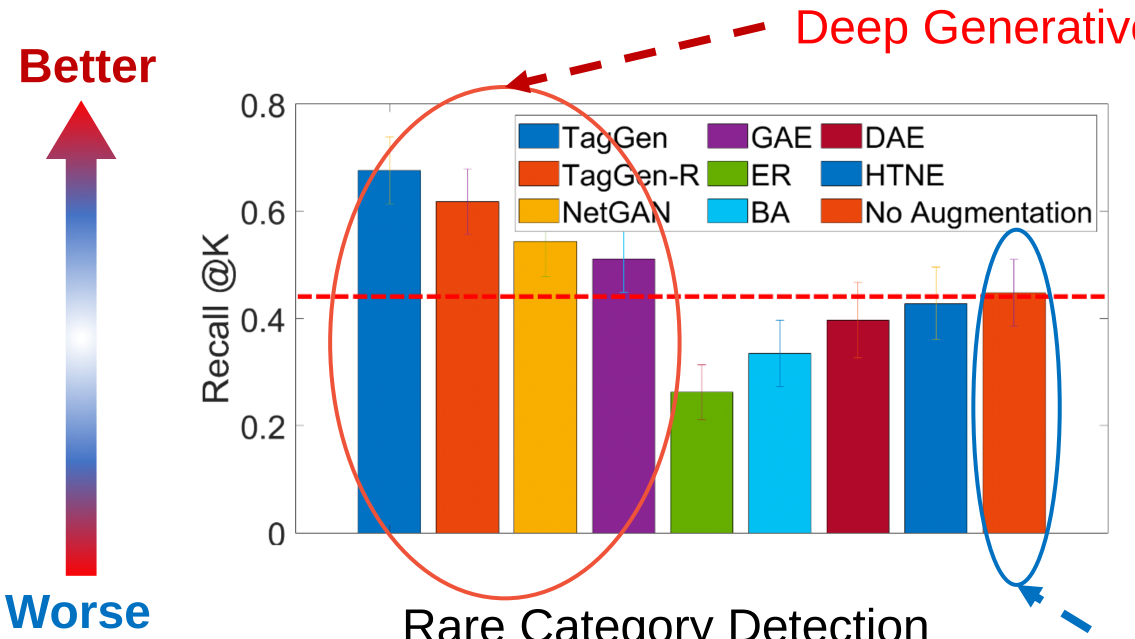
❑ S4: Graph Assembling

- **Goal:** assemble all the generated temporal random walks and generate the temporal networks.
- **Solution:** assembling rules to avoid some rare temporal occurrences (i.e., with a small degree) are not sampled.
  - ➢ Sample at least one temporal edge starting from each node with probability $p(v^{t_v})$.
  - ➢ Sample at least one temporal edge at each timestamp with probability $p(e^{t_e})$.
  - ➢ Stop until the generated graph has the same edge density as the input graphs.

• Zhou, Dawei, et al. "A Data-Driven Graph Generative Model for Temporal Interaction Networks." ACM SIGKDD. 2020.

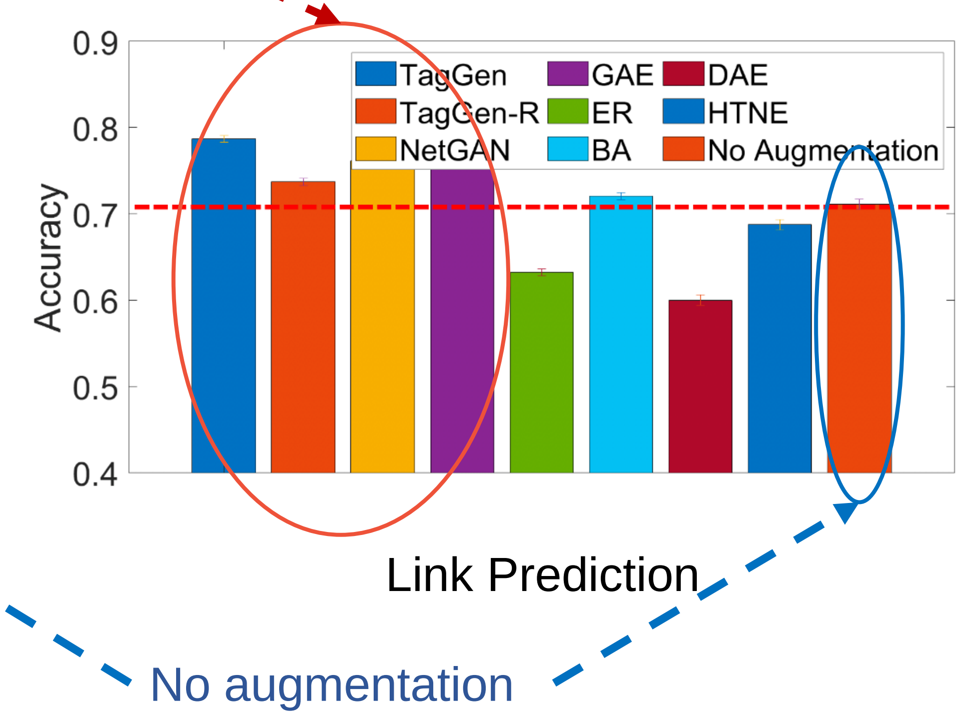
Unsupervised Graph Generation Experiment

- **Data augmentation in downstream tasks.**
  - All augmented graphs/original graph are evaluated by logistic regression.
  - 10% more edges are injected to augment the original graph.



Rare Category Detection

* K = # of examples in the target rare category.



Link Prediction

- Zhou, Dawei, et al. "A Data-Driven Graph Generative Model for Temporal Interaction Networks." ACM SIGKDD. 2020.

Rare Category Generation

Part III: Rare Category Generation

Unsupervised Rare Category Generation

- Learning data distribution of network without constraints
- Learning data distribution of temporal network without constraints

Supervised Rare Category Generation

- Learning data distribution of network with constraints

Supervised Rare Category Generation

Problem 3.4. Molecular graph generation

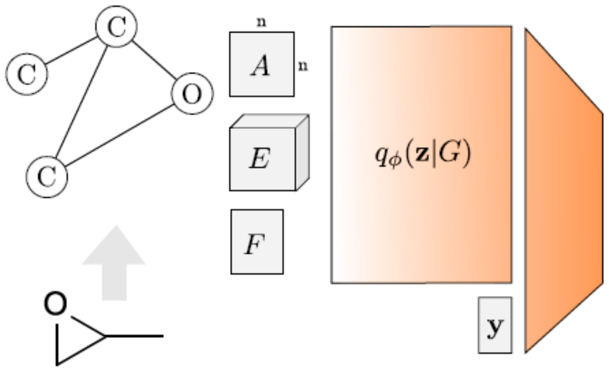
- **Given**:
  - (1) A set of molecular graphs $D = \{G_1, \dots, G_N\}$,
  - (2) Labels or constraints for graph generation,
  - (3) A rare category analysis model $f(x)$ trained in D.
- **Find**:
  - (1) A generative model $g(G)$ tailored for D,
  - (2) A set of generated graphs D' that can boost the performance of $f(x)$ in downstream rare category analysis tasks.

GraphVAE

- **Key idea:** generate a probabilistic fully-connected graph and use a graph matching algorithm to align it to the ground truth.
- **Objective function:**
  - Reconstruction loss enforces high similarity of generated graphs to the input graph G.
  - KL-divergence regularizes the space to allow for sampling of z directly from $p(z)$ instead from $q_\phi(z|G)$.

$$\mathcal{L}(\theta, \phi, G) = \underbrace{\mathbb{E}_{q_\phi(z|G)}\left[-\log p_\theta(G|z)\right]}_{\text{Reconstruction loss}} + \underbrace{D_{KL}(q_\phi(z|G)||p_\theta(z))}_{\text{Regularization}}$$

- **Encoder:**
  - Adjacency matrix A,
  - Edge attribute tensor E,
  - Node attribute matrix F,
  - Embeds the graph into continuous representation z.

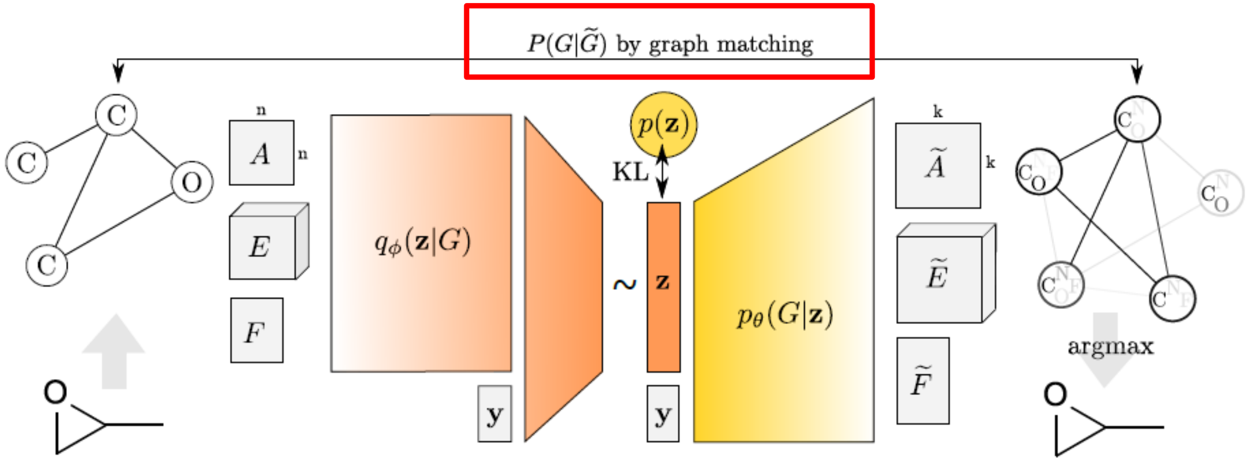


- Simonovsky, Martin et al. "Graphvae: Towards generation of small graphs using variational autoencoders." ICANN 2018.

GraphVAE

❑ Decoder:

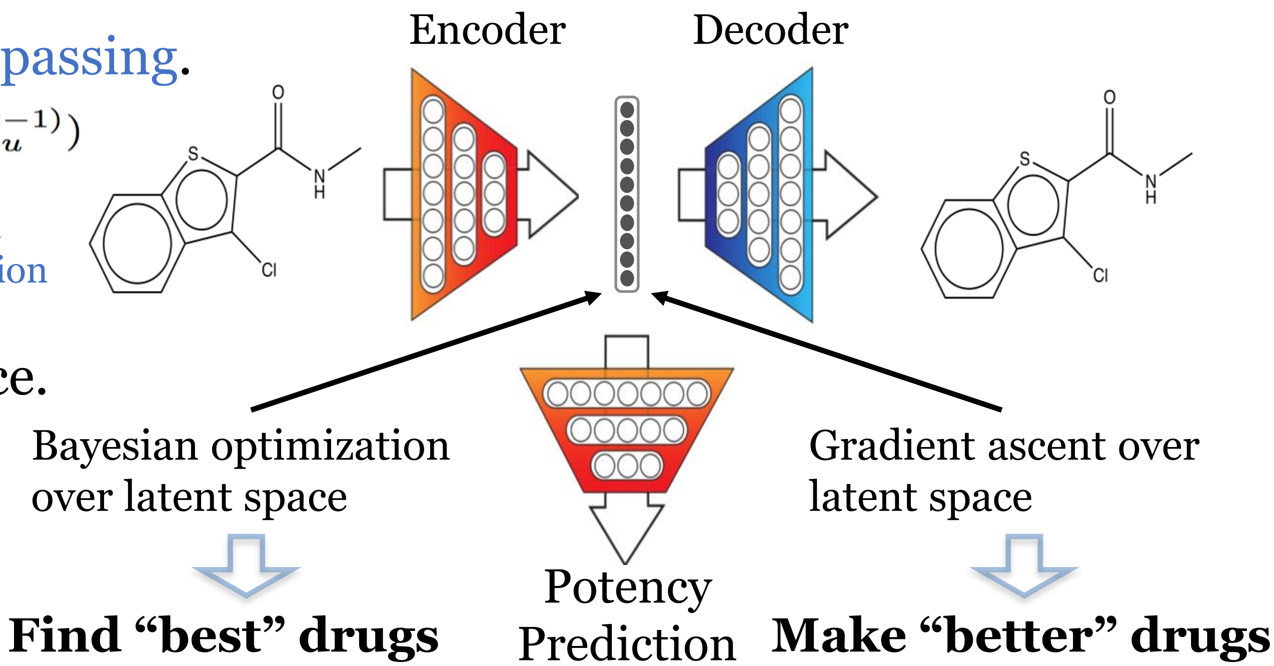
- Outputs a probabilistic fully-connected graph $\tilde{G} = (\tilde{A}, \tilde{E}, \tilde{F})$.
- $\tilde{A} \in [0,1]^{k \times k}$ contains both node probabilities $\tilde{A}_{a,a}$ and edge probabilities $\tilde{A}_{a,b}$ $(a \neq b)$.
- $\tilde{E} \in \mathbb{R}^{k \times k \times d_e}$ indicates class probabilities for edges.
- $\tilde{F} \in \mathbb{R}^{k \times d_n}$ contains class probabilities for nodes.
- Max-pooling graph matching aligns the reconstructed graph to the ground truth.
- At test time, the process can be conditioned on label y for controlled sampling.



• Simonovsky, Martin et al. "Graphvae: Towards generation of small graphs using variational autoencoders." ICANN 2018.

JT-VAE

- **Key idea:** consider each subgraph as a node, generate a tree-structured scaffold over chemical substructures, and then combine them into a molecule.
- **Graph variational autoencoder**
  - **Aims to capture the fine-grained connectivity**
  - **Graph encoder**
    - Encodes graph structure with message passing.
    - $$\underbrace{\boldsymbol{\nu}_{uv}^{(t)}}_{\text{Messages}} = \tau(\mathbf{W}_1^g \underbrace{\mathbf{x}_u}_{\text{Node feature}} + \mathbf{W}_2^g \underbrace{\mathbf{x}_{uv}}_{\text{Edge feature}} + \mathbf{W}_3^g \underbrace{\sum_{w \in N(u) \setminus v} \boldsymbol{\nu}_{wu}^{(t-1)}}_{\text{Messages from previous iteration}})$$
  - **Graph decoder**
    - Reconstructs molecule from latent space.
  - **Classifier**
    - Predicts the potency of a molecule.

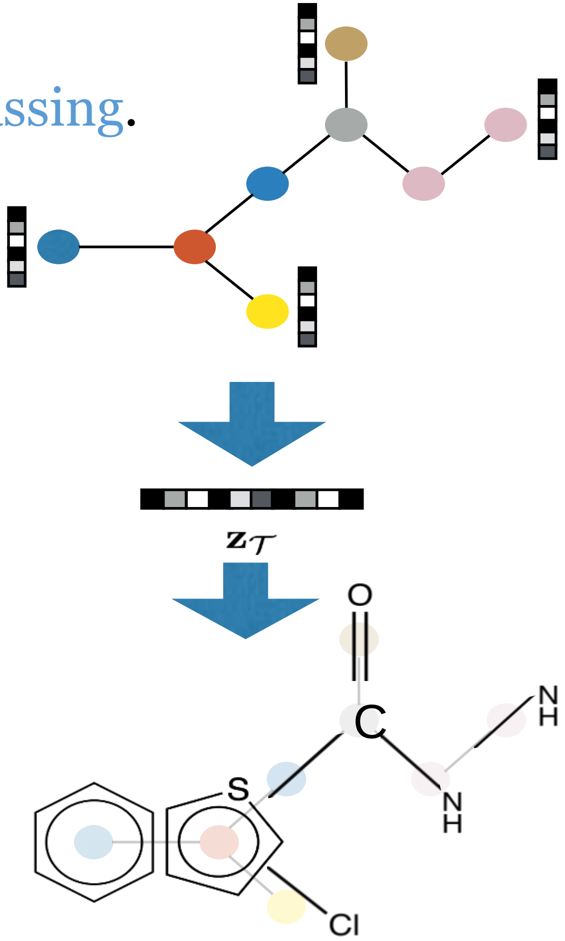


- Jin, Wengong, et al. "Junction tree variational autoencoder for molecular graph generation." ICML 2018.

JT-VAE

❑ Junction tree graph generation

- **Aims to capture the tree structure and clusters**
- **Tree encoder**
  - ➢ Vectorizes the junction tree into latent space z_T with message passing.
  - ➢ $m_{ij} = GRU(x_i, \{m_{ki}\}_{k \in N(i) \setminus j})$
  - ➢ Encode longer range interactions between nodes.
- **Tree decoder**
  - ➢ Reconstruct the junction tree from latent space z_T.



❑ Assemble two substructures into a coherent molecular graph

- Jin, Wengong, et al. "Junction tree variational autoencoder for molecular graph generation." ICML 2018.

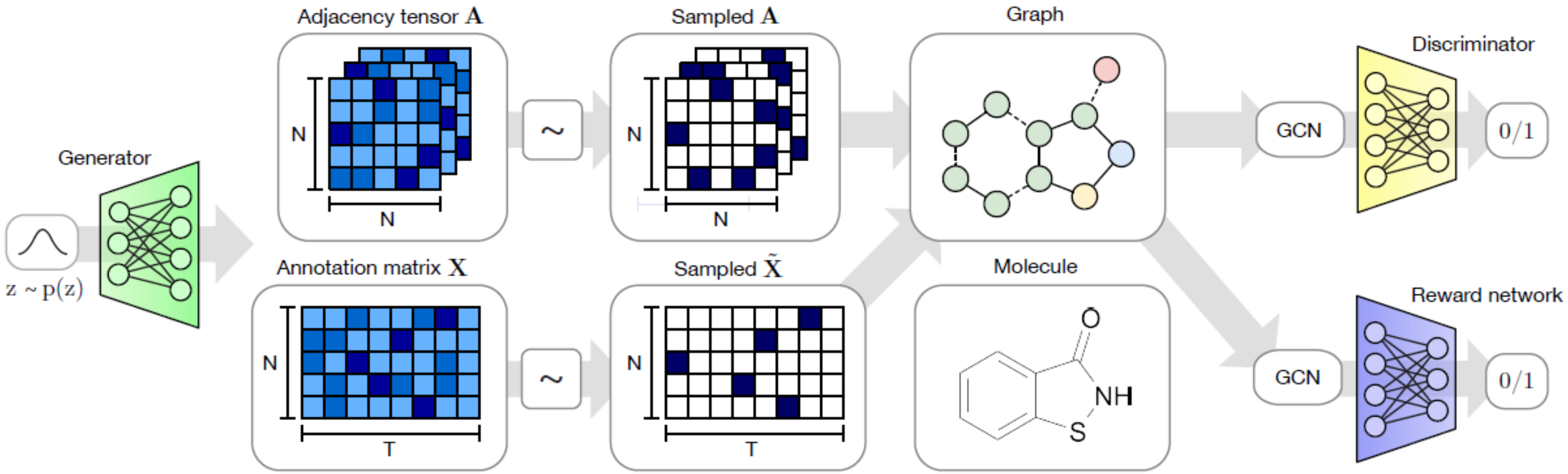
MolGAN

- **A likelihood-free molecular graph generative model**
  - **Deterministic policy gradient** is utilized to achieve useful property (e.g., to be easily synthesizable).
- **Objective function:**

$$L(\theta) = \lambda L_{WGAN} + (1 - \lambda) L_{RL}$$

Adversarial Training Loss

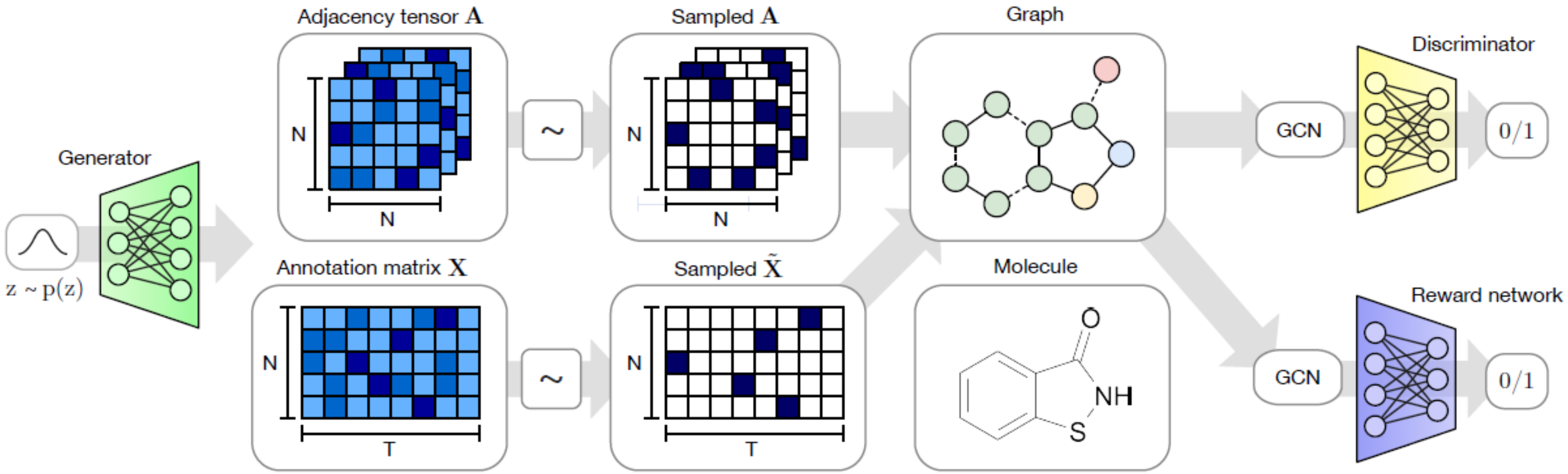
RL Loss



- De Cao, Nicola, et al. "MolGAN: An implicit generative model for small molecular graphs." ICML 2018.

MolGAN

- **Generator:**
  - Generates atom types X and bond types A from a prior distribution.
- **Discriminator:**
  - Distinguishes if the graph is real or synthesized.
- **Reward network**
  - Assigns a reward to the graph provided by external software;
  - Assigns zero to invalid molecular graph.



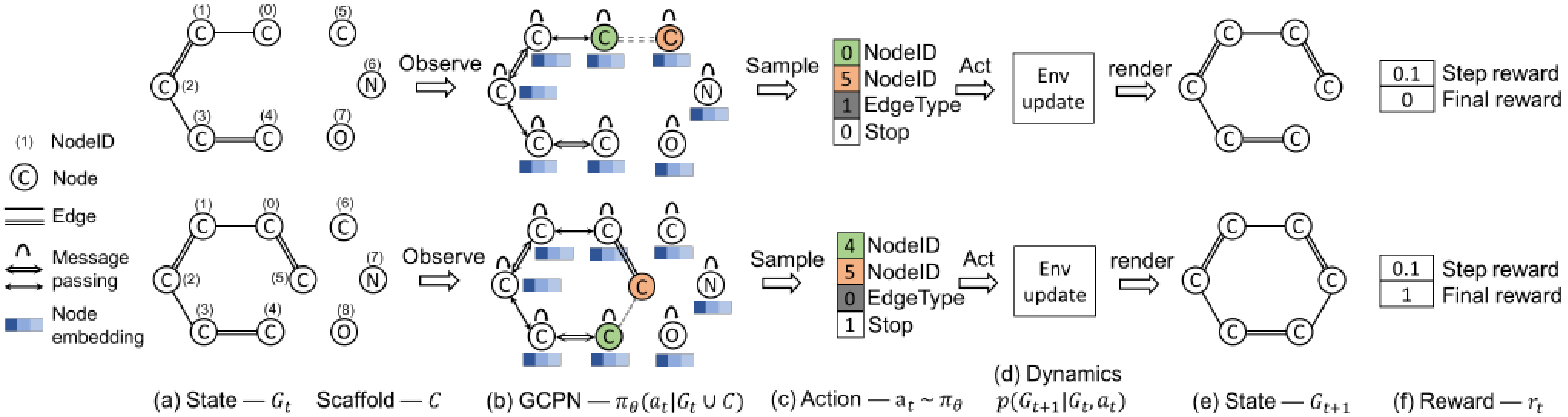
- De Cao, Nicola, et al. "MolGAN: An implicit generative model for small molecular graphs." ICML 2018.

Graph Convolutional Policy Network (GCPN)

- **Key idea:** learn a reinforcement learning agent that iteratively adds substructures and edges to the molecular graph in a chemistry-aware environment.

- **Goal:** generate graphs that maximize a given property function $S(G') \in R$.

$$\max \mathbb{E}_{G'}[S(G')]$$



- You, Jiaxuan, et al. "Graph convolutional policy network for goal-directed molecular graph generation." Advances in neural information processing systems 2018.

Supervised Graph Generation Experiment

❑ **Dataset:** ZINC250k molecule dataset.

❑ **Constrained property optimization:** focus on generating molecules with the highest possible penalized logP and QED scores.

Method	Penalized logP				QED			
	1st	2nd	3rd	Validity	1st	2nd	3rd	Validity
ZINC	4.52	4.30	4.23	100.0%	0.948	0.948	0.948	100.0%
Hill Climbing	–	–	–	–	0.838	0.814	0.814	100.0%
ORGAN	3.63	3.49	3.44	0.4%	0.896	0.824	0.820	2.2%
JT-VAE	5.30	4.93	4.49	100.0%	0.925	0.911	0.910	100.0%
GCPN	**7.98**	**7.85**	**7.80**	**100.0%**	**0.948**	**0.947**	**0.946**	**100.0%**

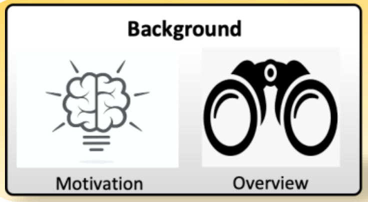
- You, Jiaxuan, et al. "Graph convolutional policy network for goal-directed molecular graph generation." Advances in neural information processing systems 2018.

Addition Algorithms

- **Unsupervised Rare Category Generation**
  - GraphGAN [Wang et al., 2018]
  - Graph-Nets [Li et al., 2018]
  - Graphite-VAE [Grover et al., 2019]
  - Misc-GAN [Zhou et al., 2019]
  - ...

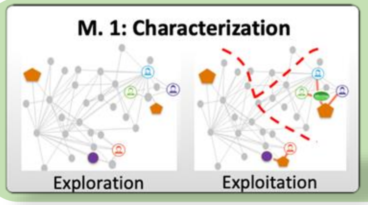
- **Supervised Rare Category Generation**
  - Reg-VAE [Ma et al., 2018]
  - MolecularRNN [Popova et al., 2019]
  - NEVAE [Samanta et al., 2019]
  - LF-MolGAN [Polsterl et al., 2019]
  - LGGAN [Fang et al,. 2019]
  - CondGEN [Yang et al., 2019]
  - ...

Roadmap



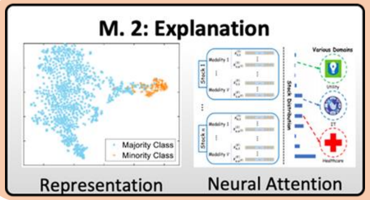
Background

- ➢ Motivation
- ➢ Research overview



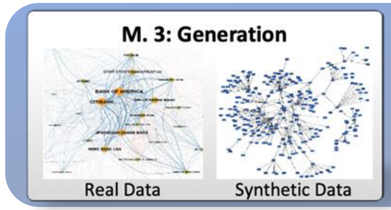
I. Rare Category Characterization

- ➢ Rare category characterization on homogeneous graphs
- ➢ Rare category characterization on heterogeneous graphs



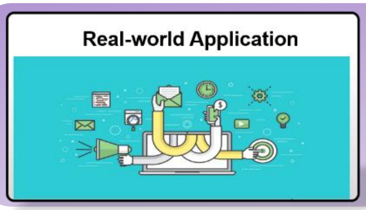
II. Rare Category Explanation

- ➢ Data insights
- ➢ Model insights



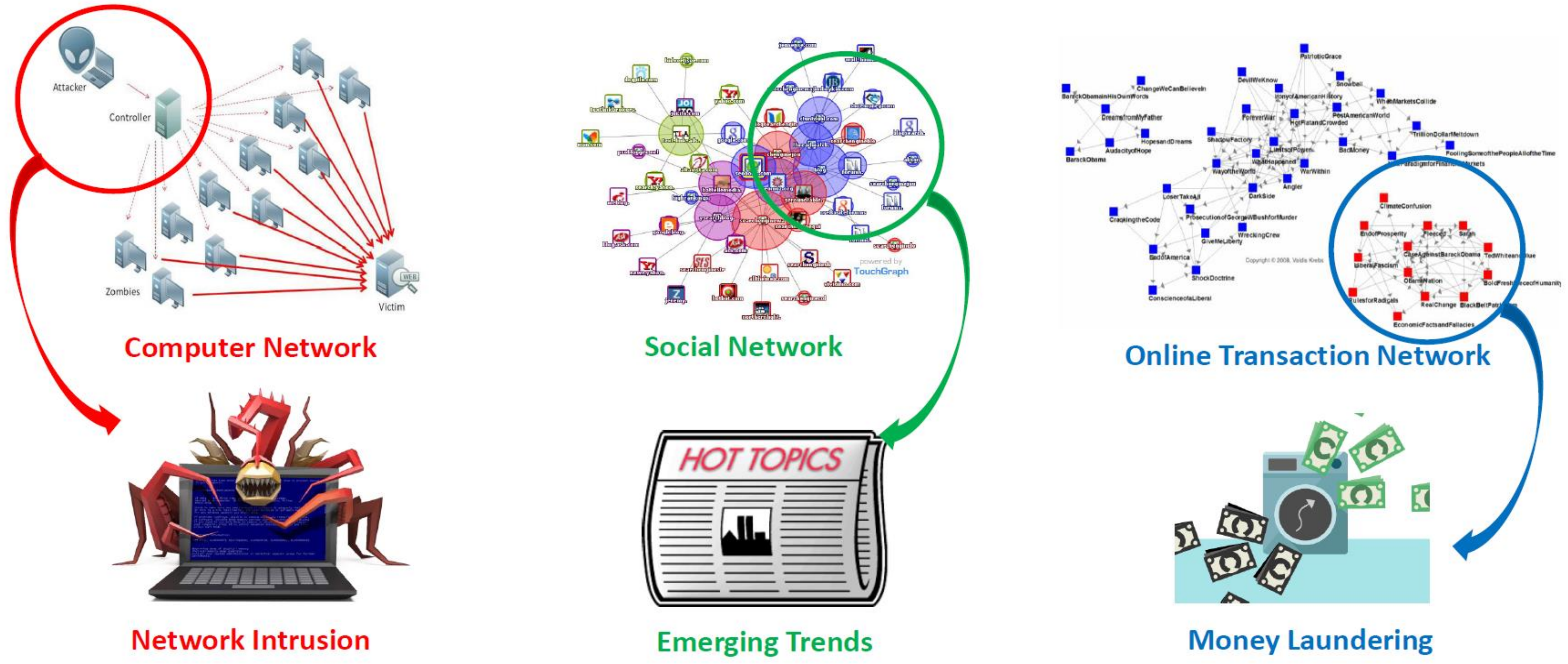
III. Rare Category Generation

- ➢ Unsupervised rare category generation
- ➢ Supervised rare category generation

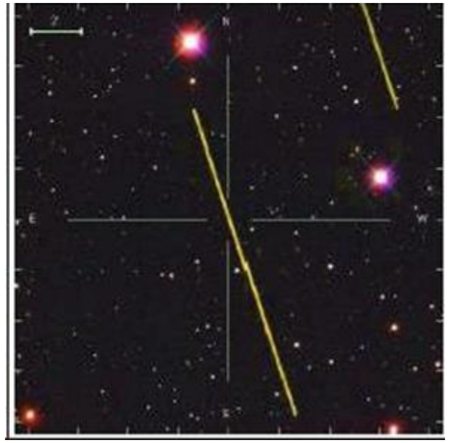


IV. Real-world Application

Real-World Applications

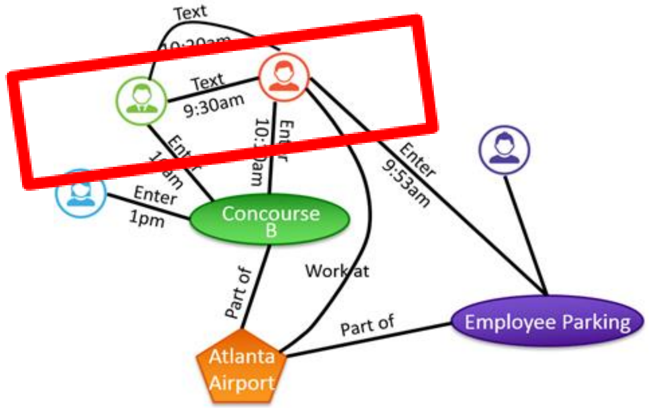


Real-World Applications



1. Out Space Discovery

2. Racial Hatred Discovery



3. Airport Insiders

4. Gene Disease

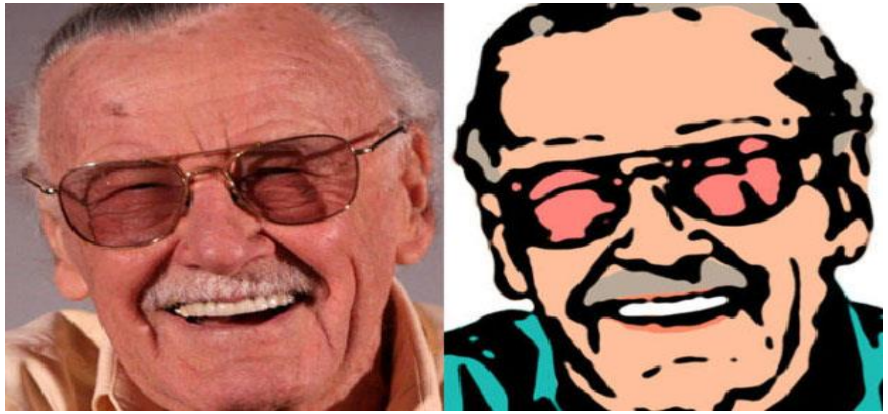
5. Identity Theft

6. Lone Wolf Terrorism

Real-World Applications

❑ Image style transfer

- GAN-based method
- Application: TikTok
- Real image → animation style



Style Transfer

❑ Graph Generation and Drug Repurposing

- Identify the similar symptoms of two diseases
- Test a therapy approved for one disease for its effect on another
- Select drugs with likelihood of success

Drug Repurposing

- Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
- Image source: https://www.linkedin.com/pulse/drug-repurposing-hopes-hurdles-abirami-santhanam-phd/

References

Introduction

- Y. X. Wang, D. Ramanan, and M. Hebert. Learning to model the tail. In Advances in Neural Information Processing Systems, 2017.
- R. Spaaij. "The enigma of lone wolf terrorism: An assessment." Studies in Conflict & Terrorism, 2010.
- E. W. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision support systems, 2010.
- D. Pelleg, and A. W. Moore: Active Learning for Anomaly and Rare-Category Detection. NIPS, 2004.
- J. He. Rare category analysis (Doctoral dissertation, Carnegie Mellon University, School of Computer Science, Machine Learning Department), 2010.
- H. He, and A. ., Edwardo. "Learning from imbalanced data." IEEE Transactions on Knowledge & Data Engineering, 2008.
- N. V. Chawla, J. Nathalie, and K. Aleksander. "Special issue on learning from imbalanced data sets." ACM Sigkdd Explorations Newsletter, 2004
- Workshop on Learning from Imbalanced Data Sets, Co-located with AAAI 2000.
- Workshop on Learning with Imbalanced Domains: Theory and Applications, Co-located with ECML/PKDD 2018.
- Workshop on Active Learning for Decision-Making from Imbalanced Observational Data, Co-located with ICML 2019.

References

Part I

- J. He, and J. Carbonell. Co-Selection of Features and Instances for Unsupervised Rare Category Analysis. SDM 2010.
- J. Sun, H. Qu, D. Chakrabarti, and C. Faloutsos. Neighborhood formation and anomaly detection in bipartite graphs. ICDM 2005.
- H. Tong, and C. Lin. Non-Negative Residual Matrix Factorization with Application to Graph Anomaly Detection. SDM 2011.
- E. A. Manzoor, S. M. Milajerdi, and L. Akoglu: Fast Memory-efficient Anomaly Detection in Streaming Heterogeneous Graphs. KDD 2016.
- D. Zhou, J. He, K. S. Candan, and H. Davulcu. MUVIR: Multi-View Rare Category Detection. IJCAI 2015.
- L. Akoglu, M. McGlohon, and C. Faloutsos. OddBall: Spotting Anomalies in Weighted Graphs. PAKDD, 2010.
- D. Zhou, S. Zhang, M. Y. Yildirim, S. Alcorn, H. Tong, H. Davulcu, and J. He: A Local Algorithm for Structure-Preserving Graph Cut. KDD 2017.
- S. Zhang, D. Zhou, M. Y. Yildirim, S. Alcorn, J. He, H. Davulcu, and H. Tong. HiDDen: Hierarchical Dense Subgraph Detection with Application to Financial Fraud Detection. SDM 2017.
- G. Pang, C. Shen, and A. van den Hengel. Deep anomaly detection with deviation networks. ACM SIGKDD 2019.
- K. Ding, J. Li, R. Bhanushali, and H. Liu. Deep anomaly detection on attributed networks. SDM 2019.

References

Part I

- V. Pavan, and W.-K. Wong. "Category detection using hierarchical mean shift." ACM SIGKDD 2009.
- K. Ding, Y. Li, J. Li, C. Liu, and H. Liu. Graph Neural Networks with High-order Feature Interactions. arXiv preprint arXiv:1908.07110.
- Z. Liu, W. Liu, P.Y. Chen, C. Zhuang, and C. Song. hpGAT: High-Order Proximity Informed Graph Attention Network. IEEE Access, 2019.
- C. Meng, S.C. Mouli, B. Ribeiro, and J. Neville. Subgraph pattern neural networks for high-order graph evolution prediction. AAAI 2018.
- A.G. Carranza, R.A. Rossi, A. Rao, and E. Koh. Higher-order spectral clustering for heterogeneous graphs. arXiv preprint arXiv:1810.02959, 2018.
- D. Zhou, J. He, H. Davulcu, and R. Maciejewski. Motif-Preserving Dynamic Local Graph Cut. BigData-2018 2018.
- E. T. Charalampos, P. Jakub, and M. Michael. Scalable Motif-aware Graph Clustering. WWW 2017.
- A.R. Benson, D.F. Gleich, and J. Leskovec. Tensor spectral clustering for partitioning higher-order network structures. SDM 2015.
- D.A. Spielman, and S.H. Teng. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. SIAM Journal on computing, 2013.

References

Part II

- D. Zhou, J. He, H. Yang, and W. Fan. SPARC: Self-Paced Network Representation for Few-Shot Rare Category Characterization. KDD 2018.
- J. Wu, J. He, and Y. Liu. ImVerde: Vertex-Diminished Random Walk for Learning Network Representation from Imbalanced Data. IEEE Bigdata, 2018.
- N. Liu, D. Shin, and X. Hu. Contextual Outlier Interpretation. IJCAI 2018.
- M., Meghanath, and L. Akoglu. "Explaining anomalies in groups with characterizing subspace rules." DMKD 2018.
- H. Lin, S. Gao, D. Gotz, F. Du, J. He, and N., Cao. Rclens: Interactive rare category exploration and identification. IEEE transactions on visualization and computer graphics, 2018.
- J. Pan, D. Han, F. Guo, D. Zhou, N. Cao, J. He, M. Xu, and W. Chen. "RCAnalyzer: visual analytics of rare categories in dynamic networks." Frontiers of Information Technology & Electronic Engineering, 2020.
- J. Kang, J. He, R. Maciejewski, and H. Tong. InFoRM: Individual Fairness on Graph Mining. ACM SIGKDD 2020.
- N. Gupta, D. Eswaran, N. Shah, L. Akoglu, and C. Faloutsos. Beyond outlier detection: Lookout for pictorial explanation. ECML 2018.
- S. Rajendran, W. Meert, V. Lenders, and S. Pollin. SAIFE: Unsupervised wireless spectrum anomaly detection with interpretable features. IEEE DySPAN 2018.

References

Part II

- J. Chen, Z. Xiu, B. Goldstein, R. Henao, L. Carin, and C. Tao. Supercharging Imbalanced Data Learning With Causal Representation Transfer. arXiv preprint arXiv:2011.12454, 2020.
- S.H. Dumpala, R. Chakraborty, S.K. Kopparapu, and T.C.S. Reseach. A Novel Data Representation for Effective Learning in Class Imbalanced Scenarios. In IJCAI 2018.
- Y. Zhao, Y. and M.K. Hryniewicki. XGBOD: improving supervised outlier detection with unsupervised representation learning. IJCNN2018.
- Z. Yang, W. Cohen, and R. Salakhudinov. Revisiting semi-supervised learning with graph embeddings. ICML 2018.
- C. Huang, Y. Li, C.C. Loy, and X. Tang. Learning deep representation for imbalanced classification. CVPR 2016.

References

Part III

- N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer. "SMOTE: synthetic minority over-sampling technique." Journal of artificial intelligence research. 2002.
- J. You, R. Ying, X. Ren, W. Hamilton, and J. Leskovec. "Graphrnn: Generating realistic graphs with deep auto-regressive models." ICML 2018.
- A. Bojchevski, O. Shchur, D. Zügner, and S. Günnemann. "Netgan: Generating graphs via random walks." ICML 2018.
- D. Zhou, L. Zheng, J. Han, and J. He. "A Data-Driven Graph Generative Model for Temporal Interaction Networks." ACM SIGKDD. 2020.
- M. Simonovsky, and N. Komodakis. "Graphvae: Towards generation of small graphs using variational autoencoders." In *International Conference on Artificial Neural Networks*, pp. 412-422. Springer, Cham, 2018.
- W. Jin, R. Barzilay, and T. Jaakkola. "Junction tree variational autoencoder for molecular graph generation." *ICML 2018.*
- N. Cao, and T. Kipf. "MolGAN: An implicit generative model for small molecular graphs." *ICML 2018*.
- J. You, B. Liu, Z. Ying, V. Pande, and J. Leskovec. "Graph convolutional policy network for goal-directed molecular graph generation." In *Advances in neural information processing systems* 2018.
- H. Wang, J. Wang, J. Wang, M. Zhao, W. Zhang, F. Zhang, X. Xie, and M. Guo, “Graphgan: Graph representation learning with generative adversarial nets.” AAAI 2018.
- D. Zhou, L. Zheng, J. Xu, and J. He, “Misc-GAN: A Multi-scale Generative Model for Graphs.” Frontiers in Big Data, 2019.

References

Part III

- Y. Li, O. Vinyals, C. Dyer, R. Pascanu, and P. Battaglia, "Learning deep generative models of graphs." arXiv preprint arXiv:1803.03324. 2018
- A. Grover, A. Zweig, and S. Ermon, "Graphite: Iterative generative modeling of graphs." In International conference on machine learning (pp. 2434-2444). PMLR 2019.
- C. Yang, P. Zhuang, W. Shi, A. Luu, and P. Li, "Conditional structure generation through graph variational generative adversarial nets." In Advances in Neural Information Processing Systems 2019.
- S. Fan, and B. Huang, "Labeled graph generative adversarial networks." arXiv preprint arXiv:1906.03220. 2019.
- T. Ma, J. Chen, and C. Xiao, "Constrained generation of semantically valid graphs via regularizing variational autoencoders." In Advances in Neural Information Processing Systems 2018.
- M. Popova, M. Shvets, J. Oliva, and O. Isayev, "MolecularRNN: Generating realistic molecular graphs with optimized properties." arXiv preprint arXiv:1905.13372. 2019.
- S. Pölsterl, and C. Wachinger, "Likelihood-Free Inference and Generation of Molecular Graphs." arXiv preprint arXiv:1905.10310. 2019
- B. Samanta, A. De, G. Jana, V. Gómez, P. Chattaraj, N. Ganguly, and M. Gomez-Rodriguez, "Nevae: A deep generative model for molecular graphs." Journal of Machine Learning Research, 2020.


Q&A