School of Computer Science
Carnegie Mellon University

Dept. of Computer Science
Rutgers University

Node and Graph Similarity: Theory and Applications

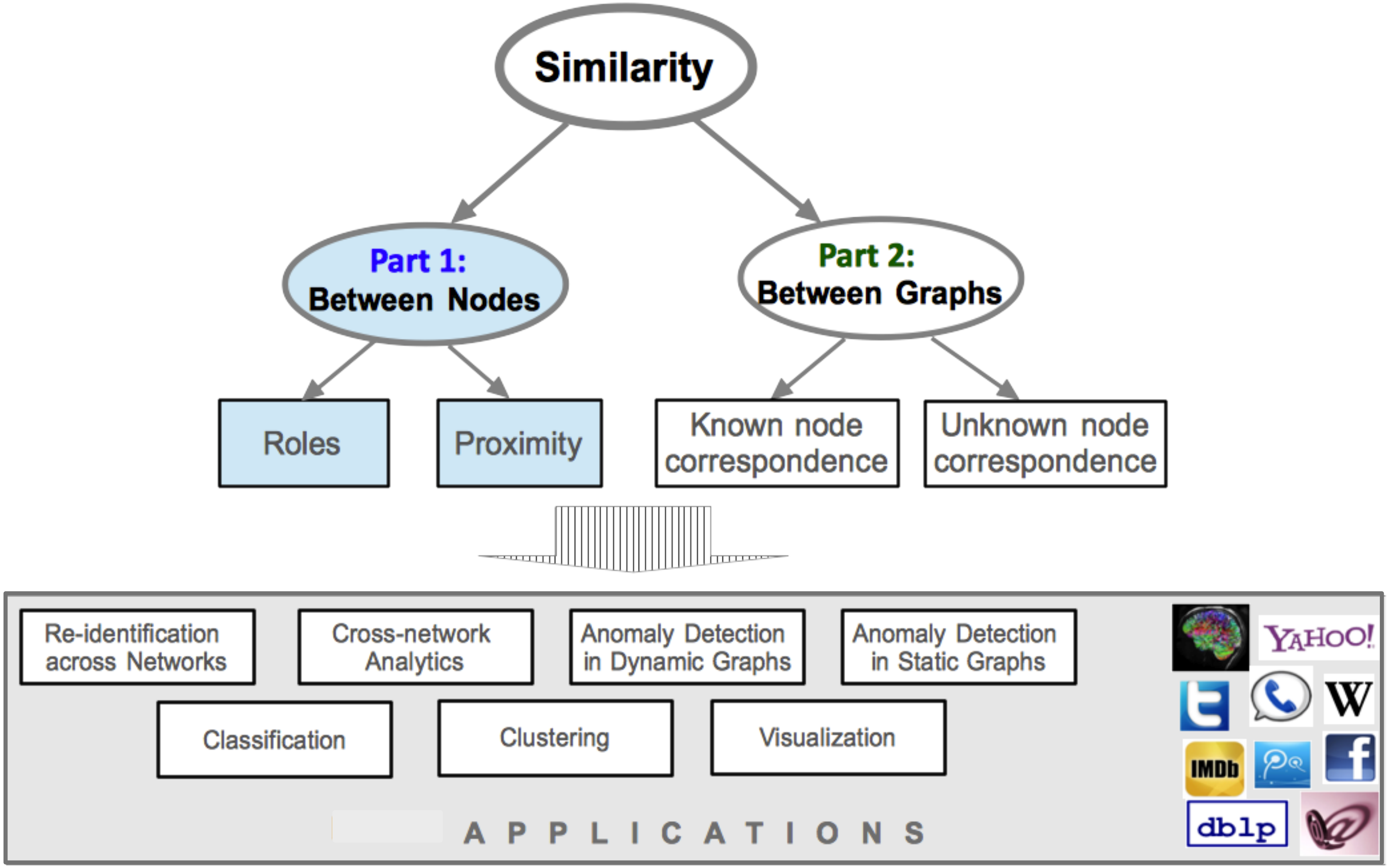
Danai Koutra (CMU)
Tina Eliassi-Rad (Rutgers)
Christos Faloutsos (CMU)

ICDM 2014, Monday December 15th 2014, Shenzhen, China

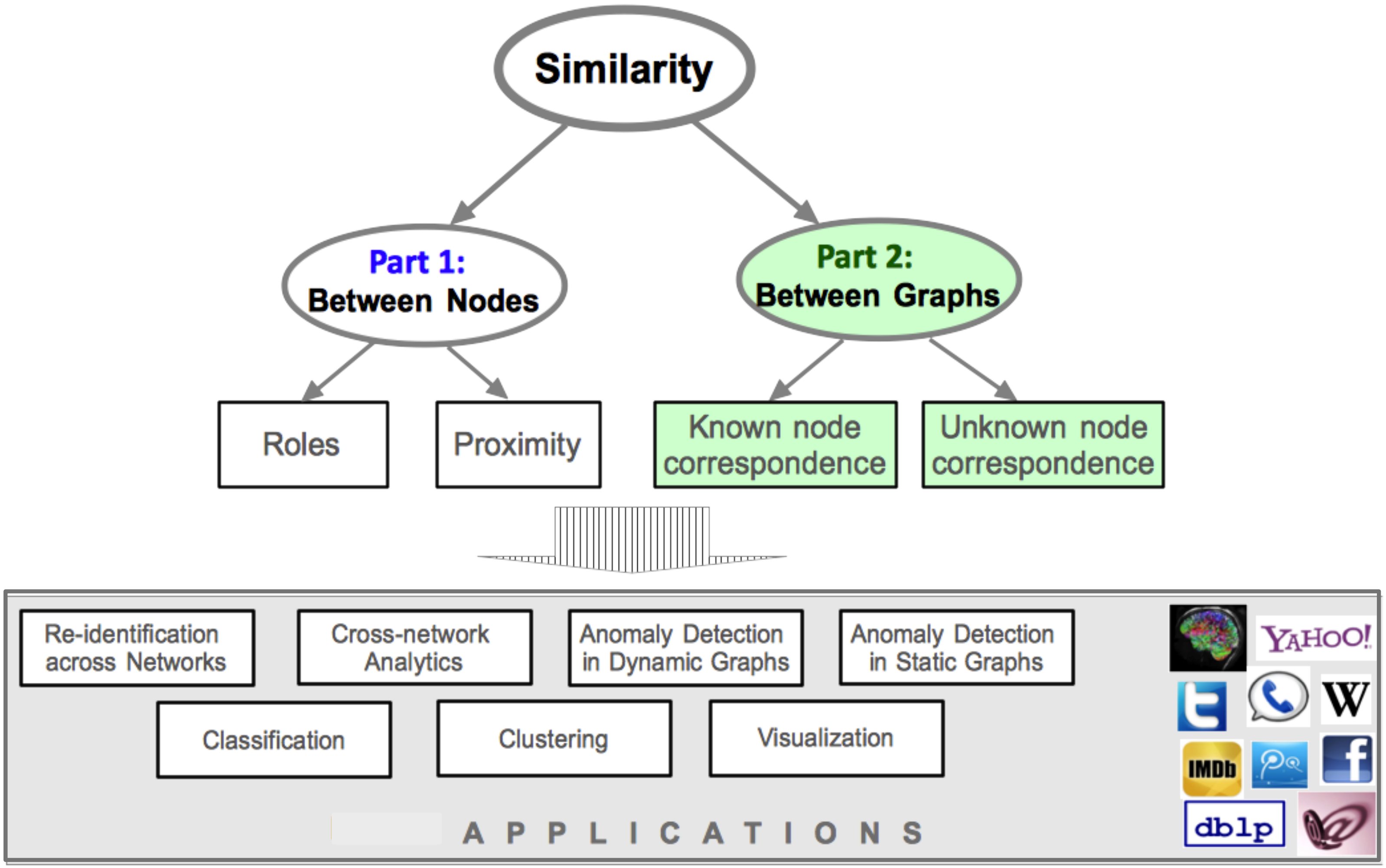
Who we are

- Danai Koutra, CMU
  - Node and graph similarity, summarization, pattern mining
  - http://www.cs.cmu.edu/~dkoutra/
- Tina Eliassi-Rad, Rutgers
  - Data mining, machine learning, big complex networks analysis
  - http://eliassi.org/
- Christos Faloutsos, CMU
  - Graph and stream mining, …
  - http://www.cs.cmu.edu/~christos

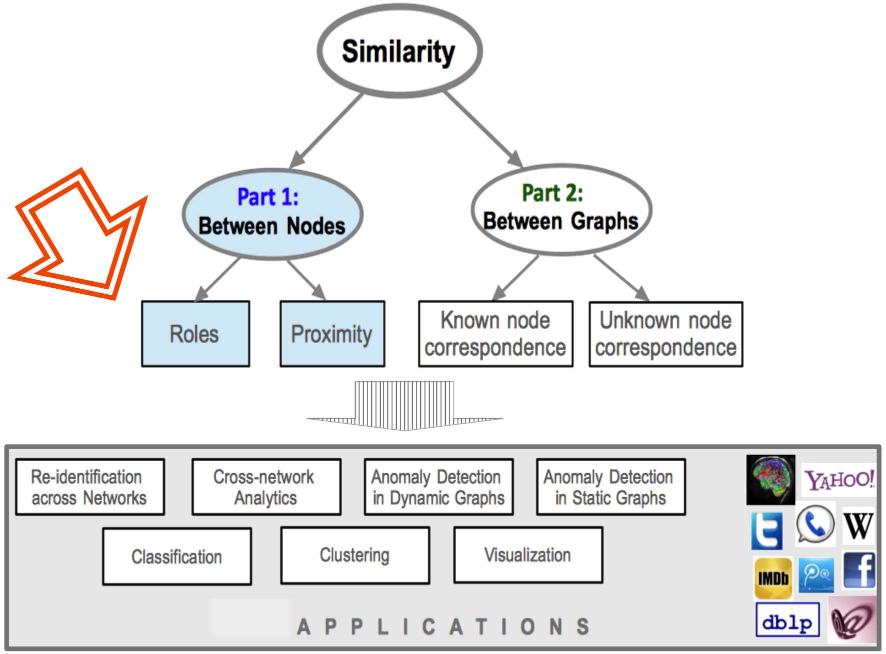
What we will cover

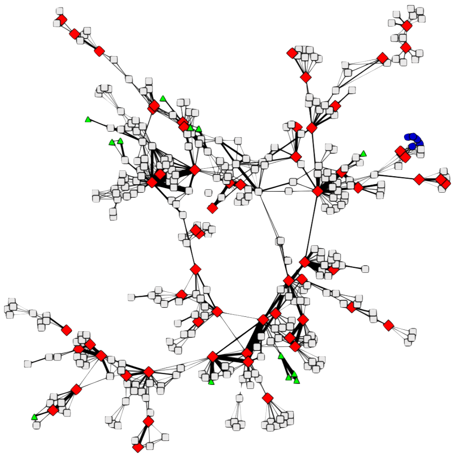


What we will cover



Part 1a
Similarity between Nodes: Roles





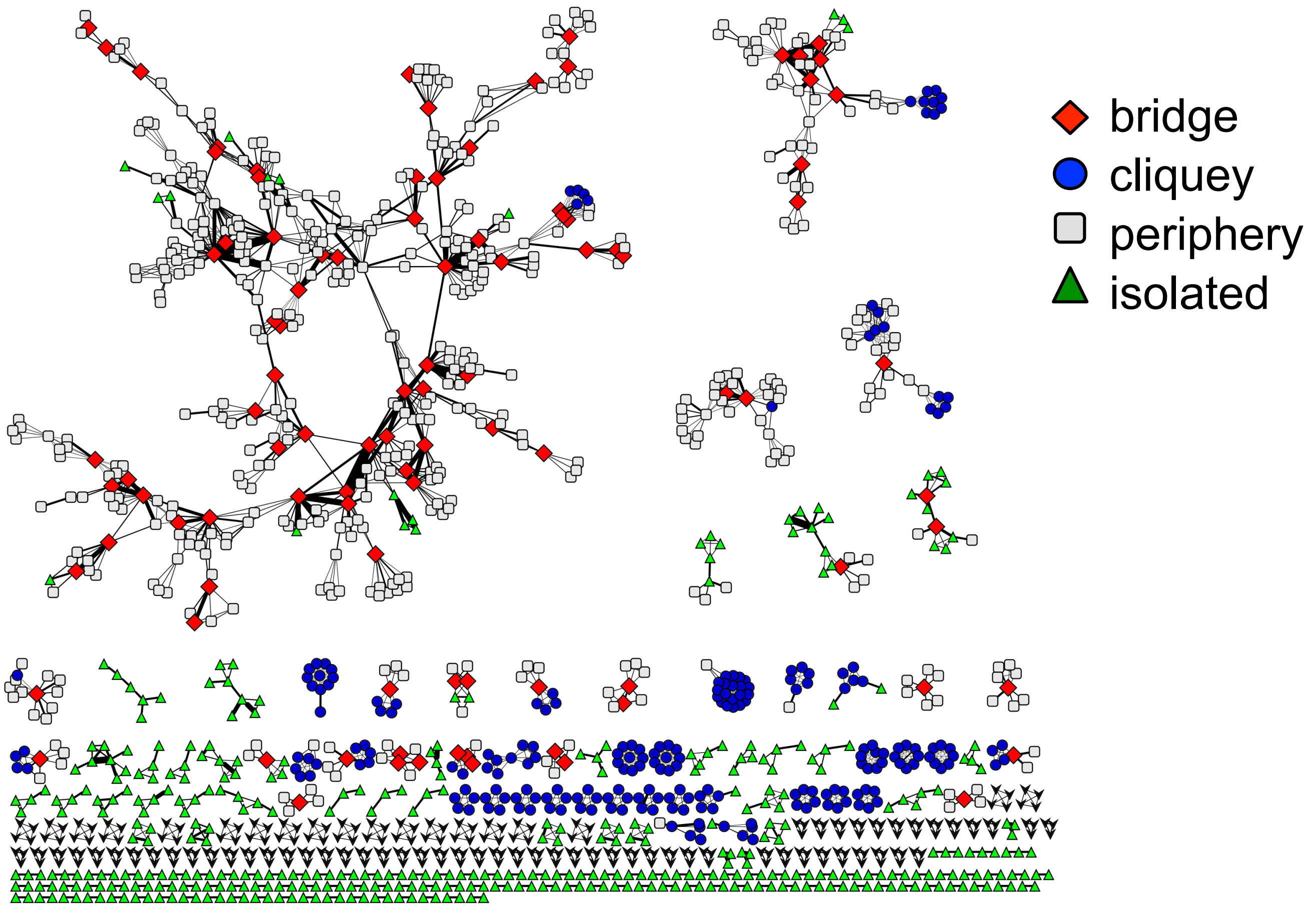
Roadmap

- Node Roles
  - What are roles
  - Roles and communities
  - Roles and equivalences (from sociology)
  - Roles (from data mining)
  - Summary
- Node Proximity (after coffee break)

What are roles?

- "Functions" of nodes in the network
  - Similar to functional roles of species in ecosystems
- Measured by **structural behaviors**
- Examples
  - centers of stars
  - members of cliques
  - peripheral nodes
  - …

Example of Roles



Network Science Co-authorship Graph
[Newman 2006]

Research Questions

- Given a graph, how can we automatically discover roles (or functions) of nodes?
- How can we make sense of these roles?
- Are there good features that we can extract for nodes that indicate role-membership?
- How are roles different from communities and from positions/equivalences (from sociology)?
- What are the applications in which these discovered roles can be effectively used?

Why are Roles Important?

- Encode complex behavior
- Map nodes into a useful lower dimensional space
  - Easier to do node similarity there
- Generalize across networks
- Have many useful applications

Applications of Role Discovery

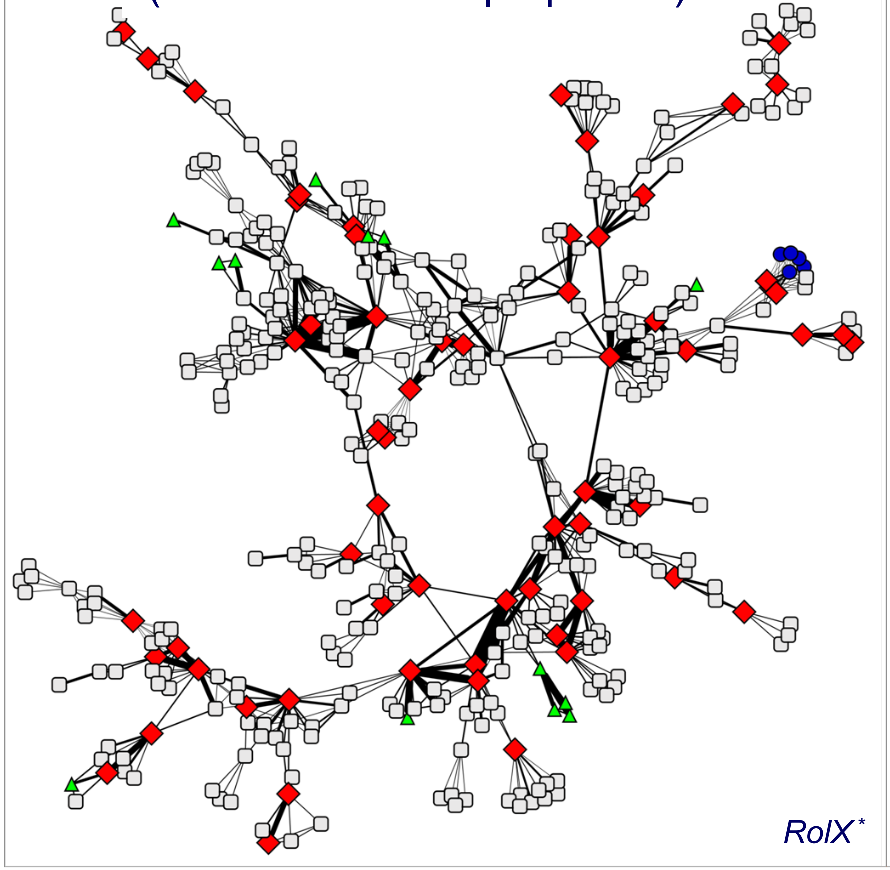
Task	Use Case
Role query	Identify individuals with similar behavior to a known target
Role outliers	Identify individuals with unusual behavior
Role dynamics	Identify unusual changes in behavior
Re-identification	Identify individuals in an anonymized network
Role transfer	Use knowledge of one network to make predictions in another
Network similarity	Determine network compatibility for knowledge transfer
Information dissemination	Cut *k* edges to minimize dissemination on a network
…	…

Roadmap

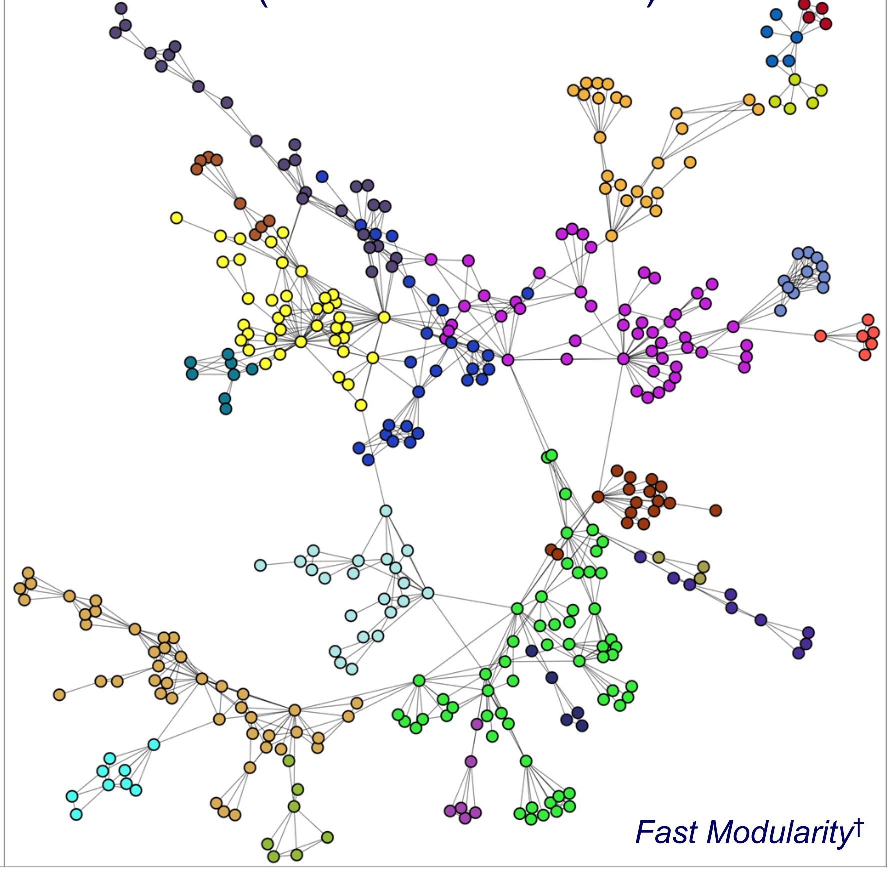
- Node Roles
  - What are roles
  - Roles and communities
  - Roles and equivalences (from sociology)
  - Roles (from data mining)
  - Summary
- Node Proximity

Roles and Communities are Complementary

<u>Roles</u>
(similar structural properties)



<u>Communities</u>
(well-connectedness)



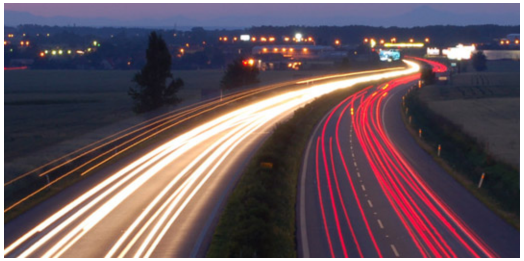
* Henderson, et al. 2012; † Clauset, et al. 2004

Roles and Communities

Consider the social network of a CS dept

- Roles
  - Faculty
  - Staff
  - Students
  - …
- Communities
  - AI lab
  - Database lab
  - Architecture lab
  - …

Roadmap



- Node Roles
  - What are roles
  - Roles and communities
  - Roles and equivalences (from sociology)
  - Roles (from data mining)
  - Summary
- Node Proximity

Roles == Positions in Sociology

- Two nodes that have the same position are in an *equivalence relation*
- Equivalence, *E*, is any relation that satisfies these 3 conditions:

1. *Transitivity*:

$$(a, b), (b, c) \in E \Rightarrow (a,c) \in E$$

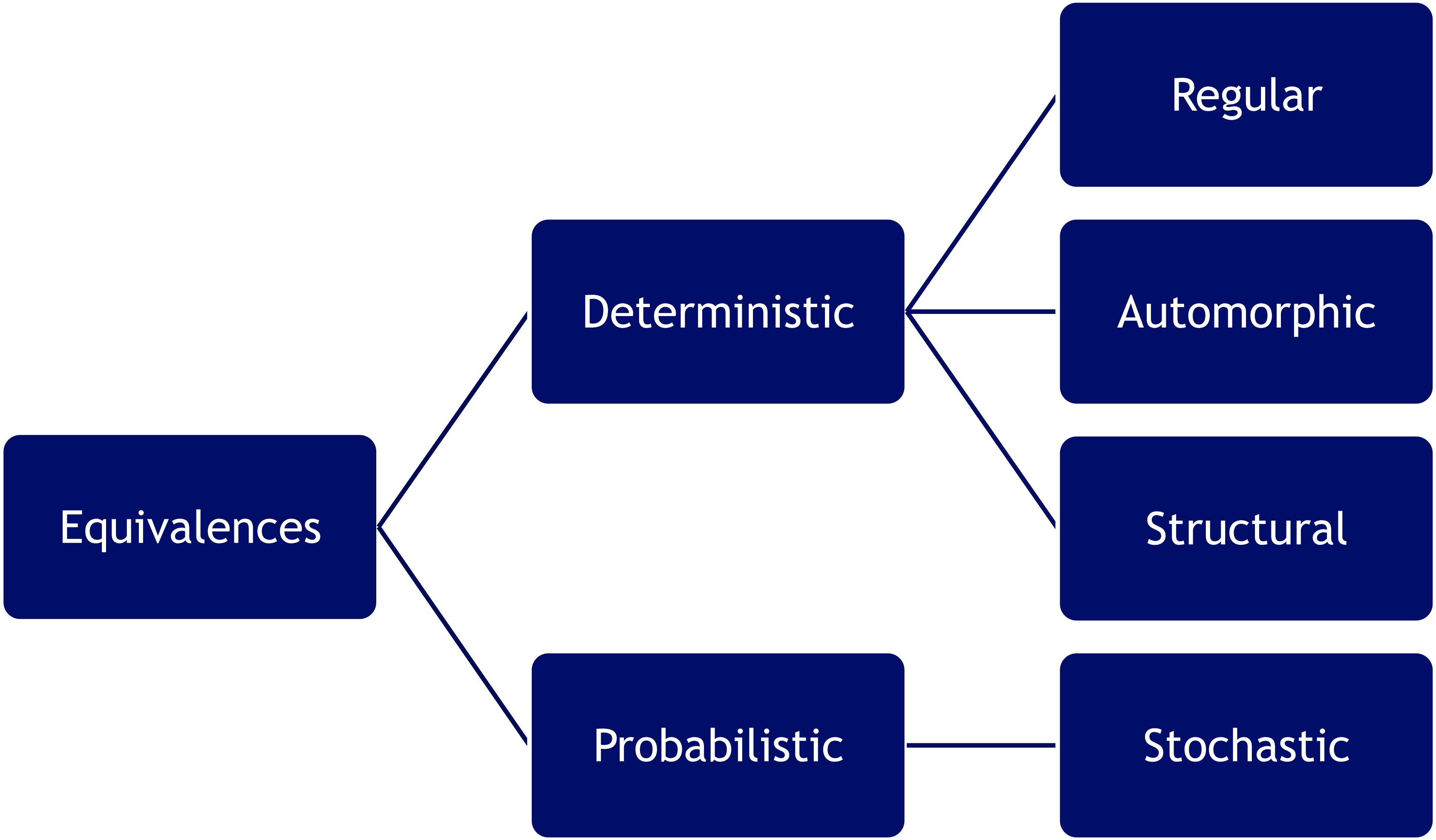
2. *Symmetry*:

$$(a, b) \in E \text{ iff } (b, a) \in E$$

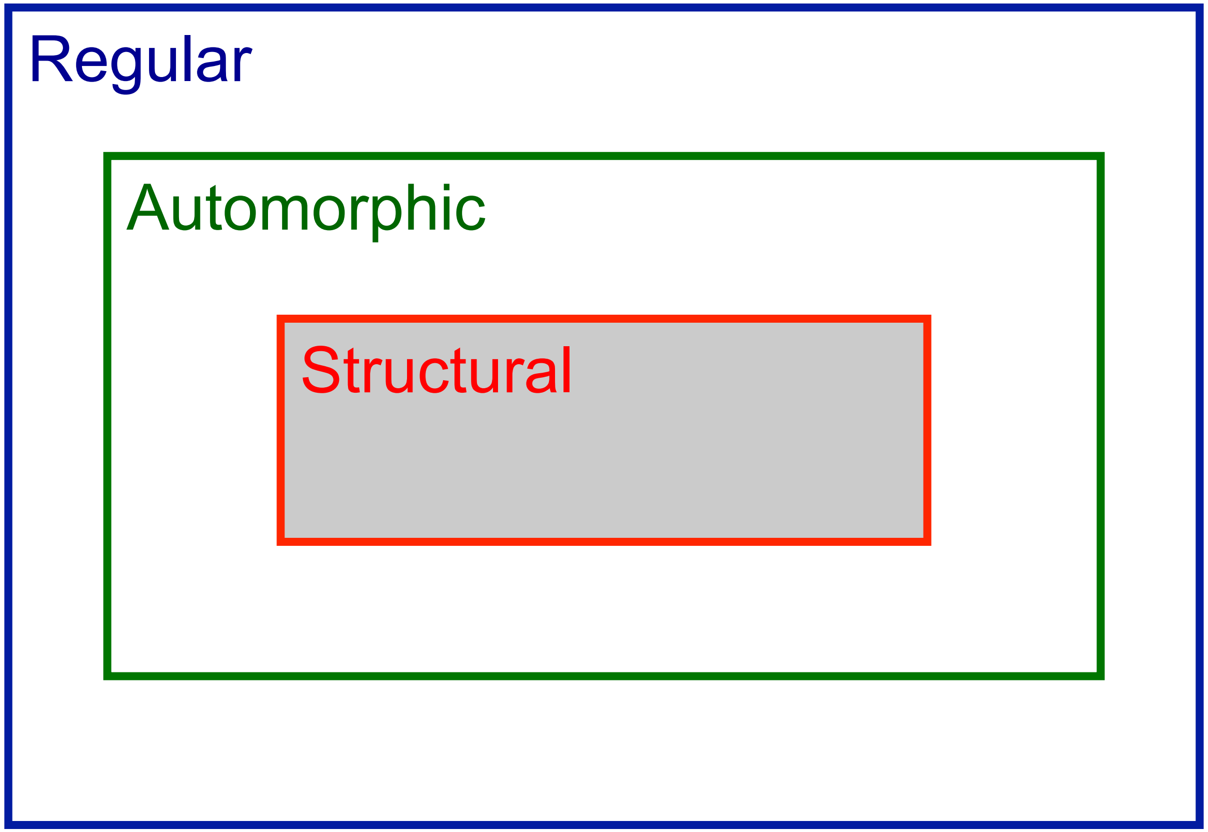
3. *Reflexivity*:

$$(a, a) \in E$$

Equivalences

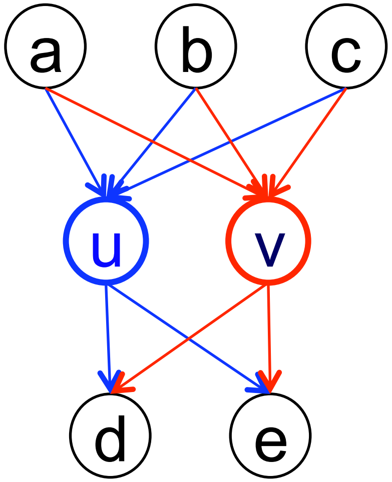


Deterministic Equivalences

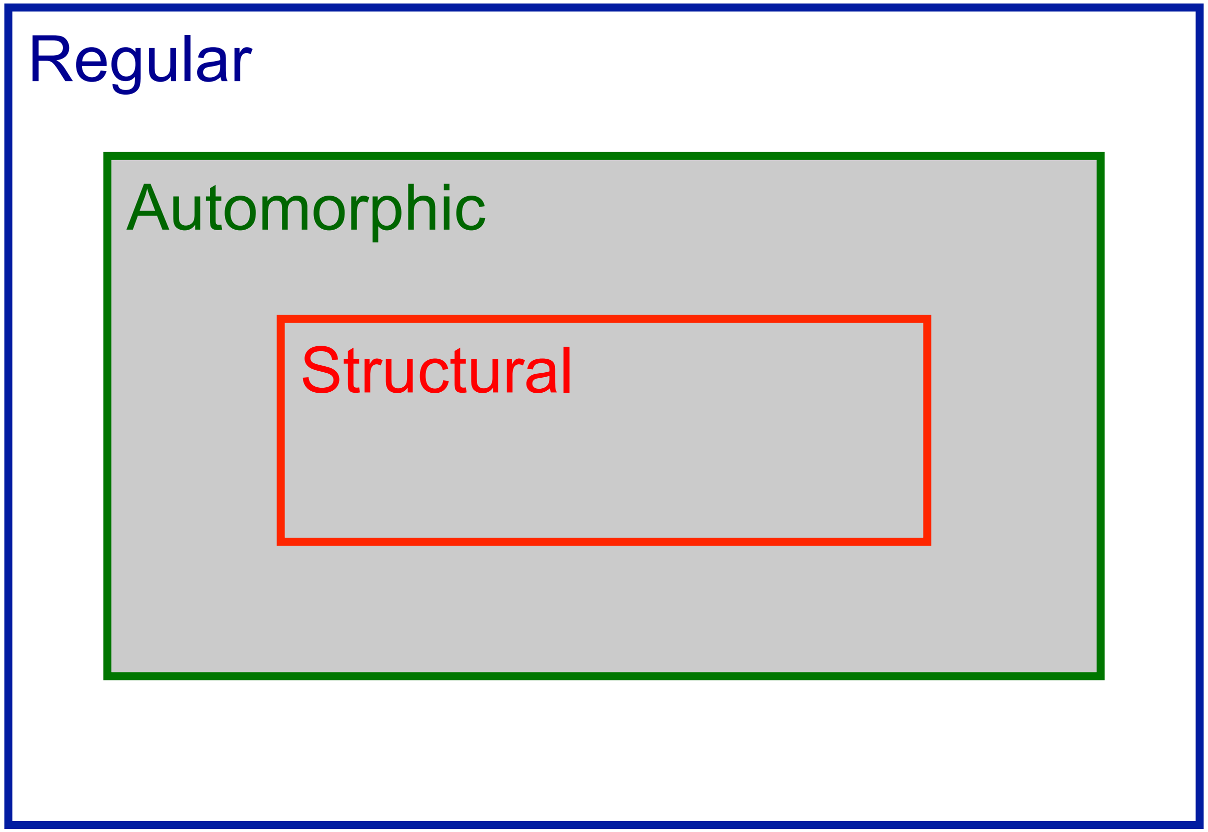


Structural Equivalence

- [Lorrain & White, 1971]
- Two nodes u and v are structurally equivalent if they have the same relationships to all other nodes
- Hypothesis: Structurally equivalent nodes are likely to be similar in other ways – i.e., you are your friend
- Weights & timing issues are not considered
- Rarely appears in real-world networks

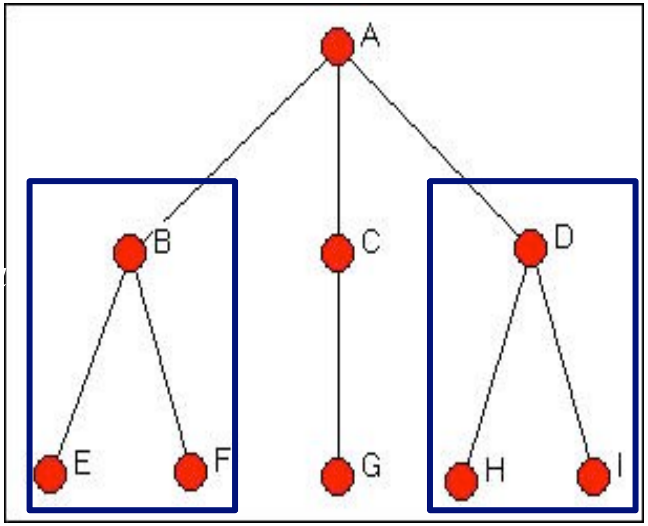


Deterministic Equivalences

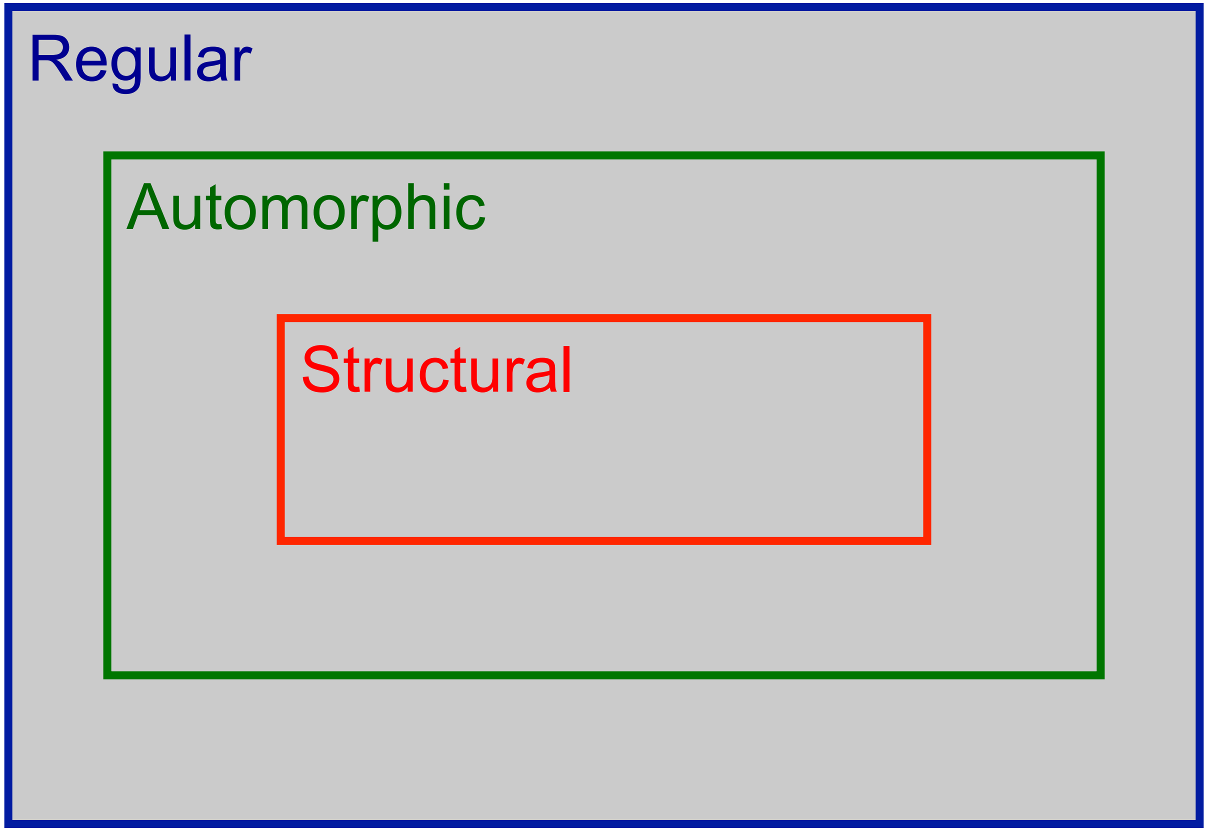


Automorphic Equivalence

- [Borgatti, et al. 1992; Sparrow 1993]
- Two nodes *u* and *v* are automorphically equivalent if all the nodes can be relabeled to form an isomorphic graph with the labels of *u* and *v* interchanged
  - Swapping *u* and *v* (possibly along with their neighbors) does not change graph distances
- Two nodes that are automorphically equivalent share exactly the same label-independent properties

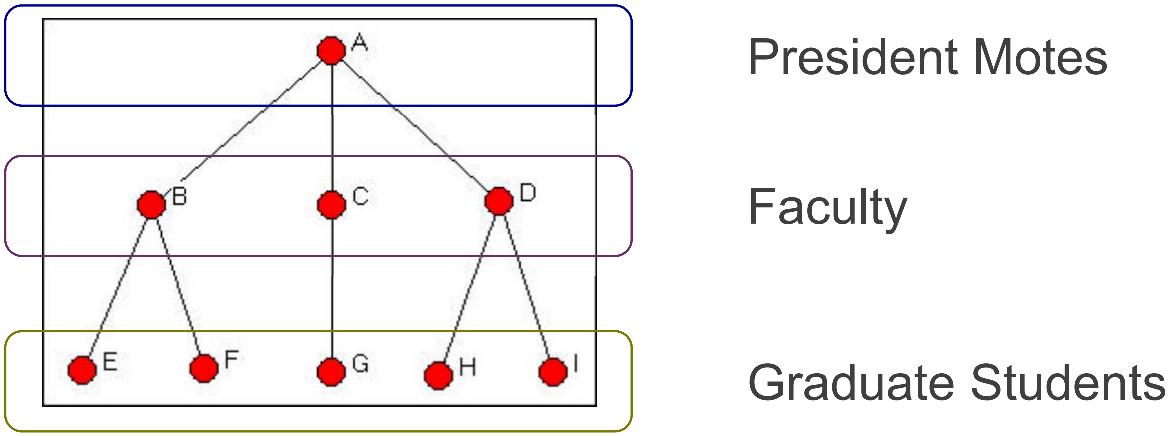


Deterministic Equivalences



Regular Equivalence

- [Everett & Borgatti, 1992]
- Two nodes *u* and *v* are regularly equivalent if they are equally related to equivalent others

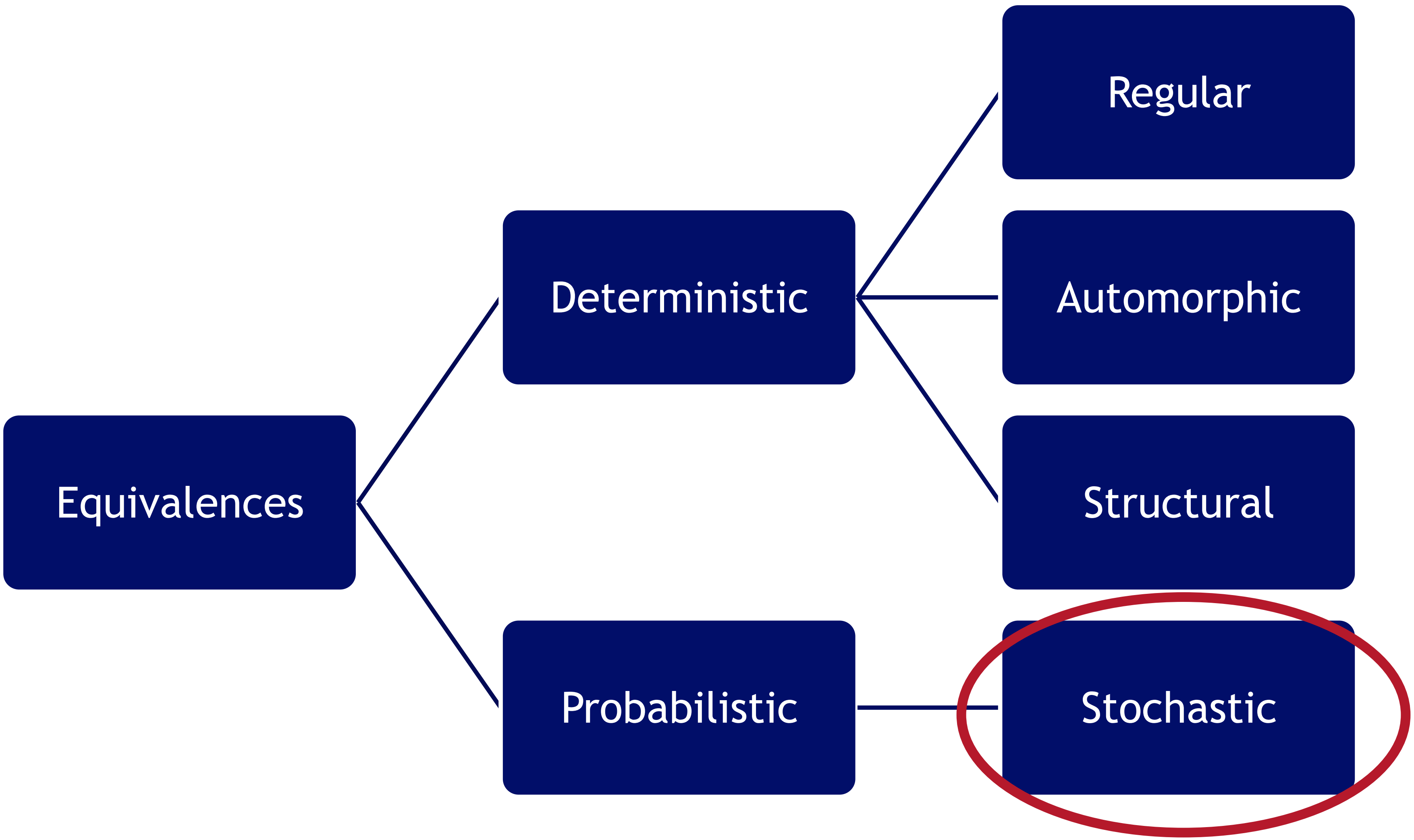


Hanneman, Robert A. and Mark Riddle. 2005. Introduction to social network methods. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/~hanneman/)

Examples of Algorithms for Deterministic Equivalences

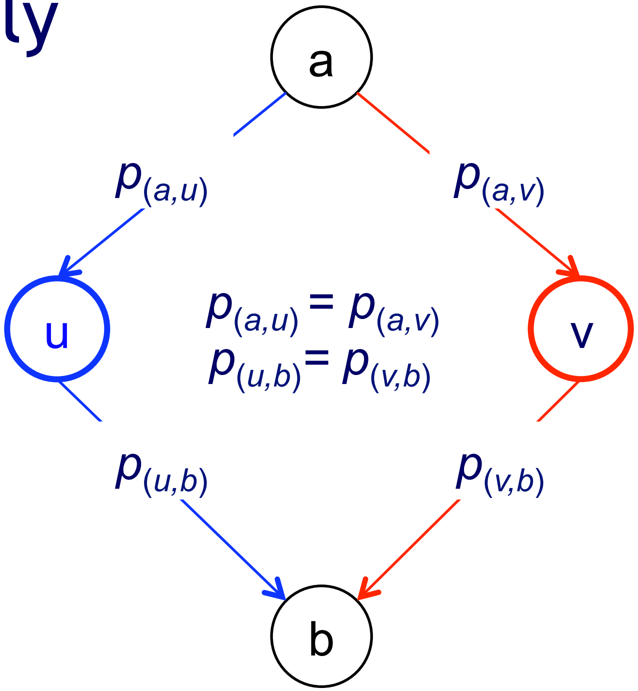
- **Structural equivalence**
  - CONCOR (CONvergence of iterated CORrelations) [Breiger et al. 1975]
  - STRUCUTRE [Burt 1976]
  - Numerical optimization with tabu search [UCINET]
  - Local optimization [Pajek]
- **Automorphic equivalence**
  - Use numerical signatures on degree sequences of neighborhoods [Sparrow, 1993]; scales linearly with the number of edges
- **Regular equivalence**
  - Maximal Regular coloration [Everett & Borgatti, 1997]; a polynomial time algorithm

Equivalences



Stochastic Equivalence

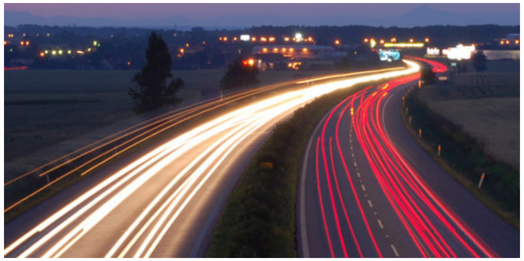
- [Holland, et al. 1983; Wasserman & Anderson, 1987]
- Two nodes are stochastically equivalent if they are "exchangeable" w.r.t. a probability distribution
- Similar to structural equivalence but probabilistic



Examples of Algorithms for Stochastic Equivalence

- Many algorithms exist here
- Most recent approaches are generative [Airoldi, et al 2008]
- Some choice points
  - Single [Kemp, et al 2006] vs. mixed-membership [Koutsourelakis & Eliassi-Rad, 2008] positions
  - Parametric vs. non-parametric models

Roadmap



- Node Roles
  - What are roles
  - Roles and communities
  - Roles and equivalences (from sociology)
  - Roles (from data mining)
  - Summary
- Node Proximity

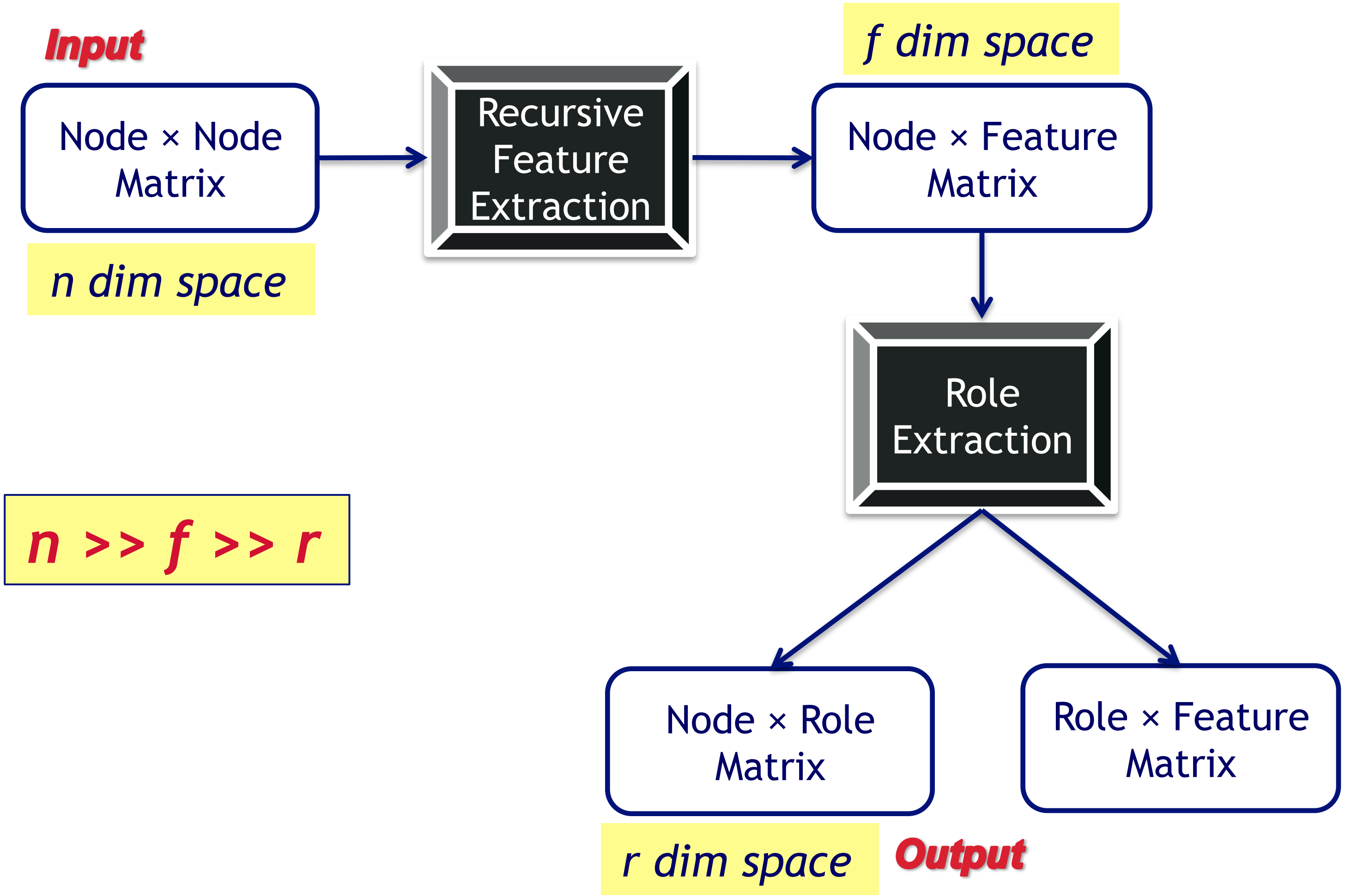
Algorithms for Role Discovery

Algorithms	What they do?	Publication
ReFeX	Recursive Feature Extraction	[Henderson *et al.*, KDD 2011]
RolX	Role Extraction from Structural Features	[Henderson *et al.*, KDD 2012]
GLRD	Guided Learning for Role Discovery	[Gilpin *et al.*, KDD 2013]
DBMM	Dynamic Behavioral Mixed-membership Model	[Rossi *et al.*, WSDM 2013]
MRD	Multi-relational Role Discovery	[Gilpin *et al.*, in preparation]
LearnMelt	Learning to Minimize Information Dissemination	[Le *et al.*, in preparation]
…		

RolX: Role eXtraction

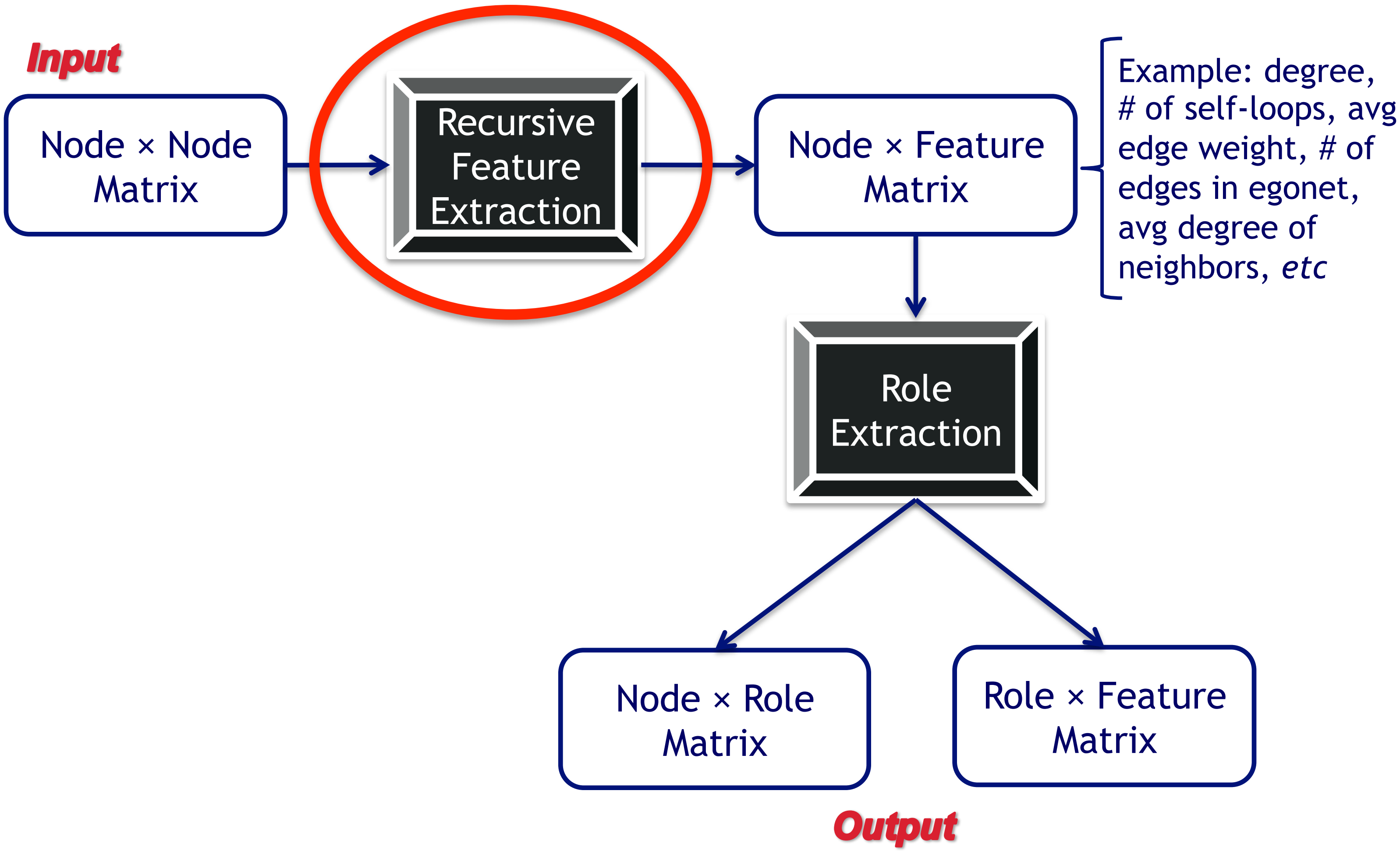
- Introduced by Henderson et al. KDD 2012
- Automatically extracts the underlying roles in a network
  - No prior knowledge required
- Determines the number of roles automatically
- Assigns a mixed-membership of roles to each node
- Scales linearly on the number of edges

RolX: Flowchart

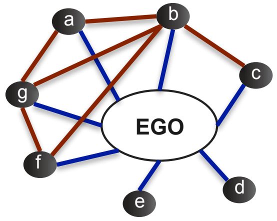


$n >> f >> r$

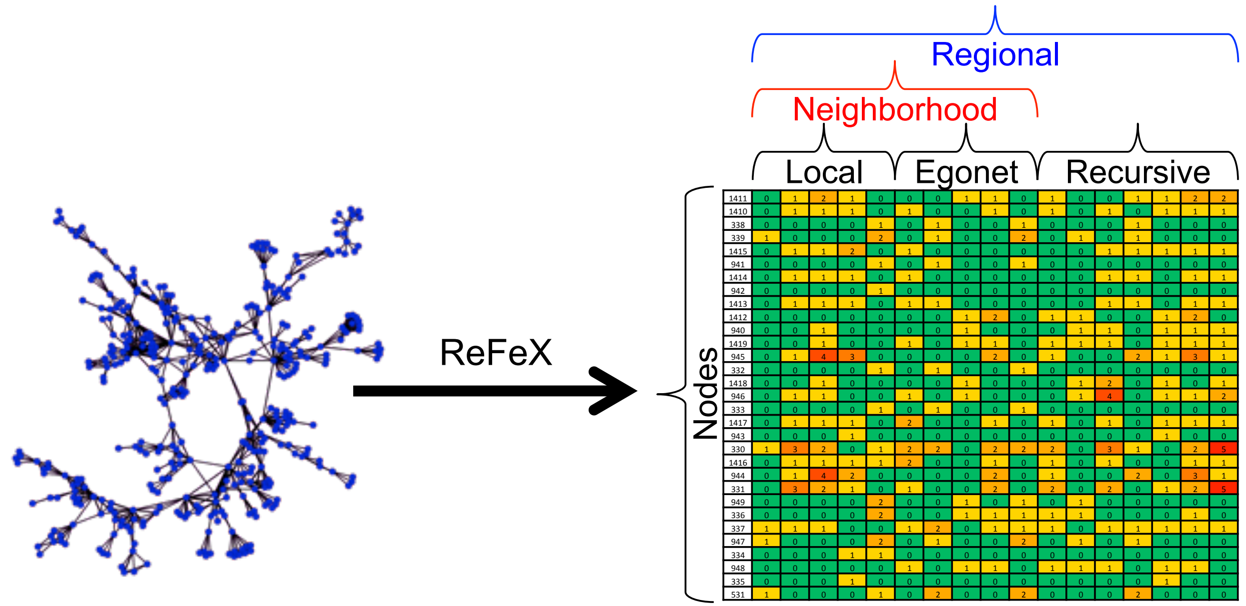
RolX: Flowchart



Recursive Feature Extraction



- ReFeX [Henderson, et al. 2011a] turns network connectivity into recursive structural features



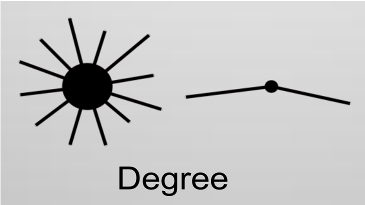
- Neighborhood features: What is your connectivity pattern?
- Recursive Features: To what kinds of nodes are you connected?

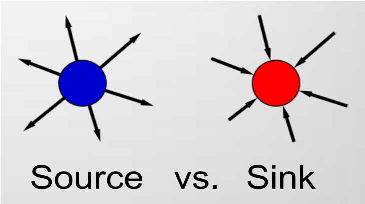
ReFeX: Structural Features

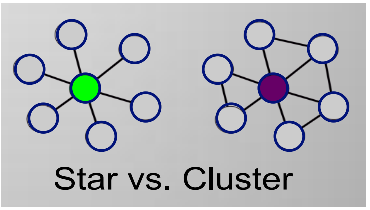
Regional

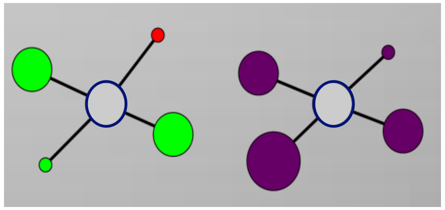
Neighborhood

- Local
  - Essentially measures of the node degree
- Egonet
  - Computed based on each node’s ego network
  - Examples
    - # of within-egonet edges
    - # of edges entering & leaving the egonet
- Recursive
  - Some aggregate (mean, sum, max, min, …) of another feature over a node’s neighbors
  - Aggregation can be computed over any real-valued feature, including other recursive features

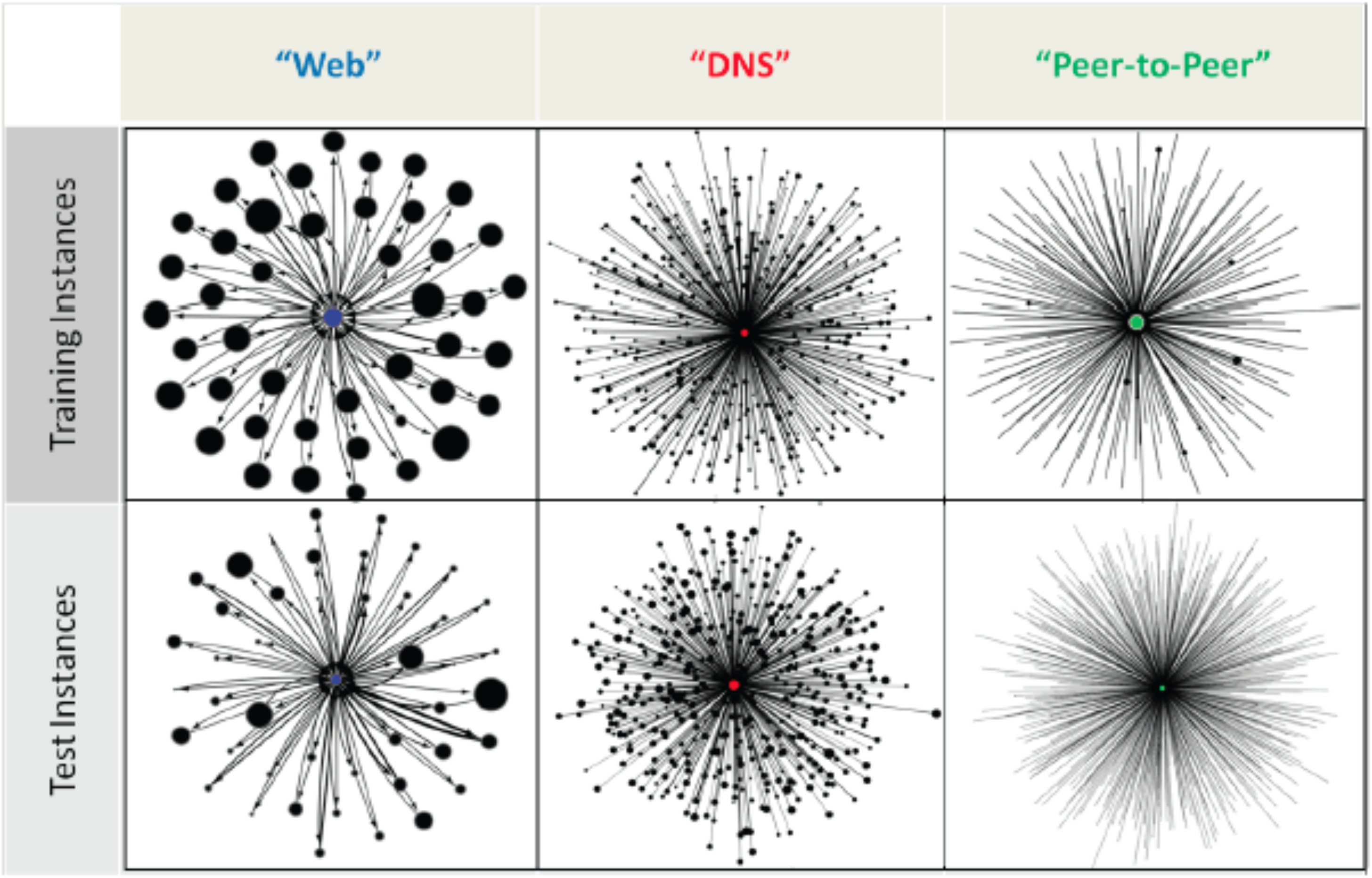








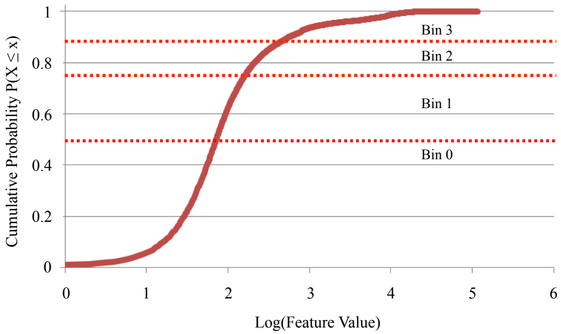
ReFex Intuition: Regional Structure Matters



Node sizes indicate communication volume relative to the central node in each frame.

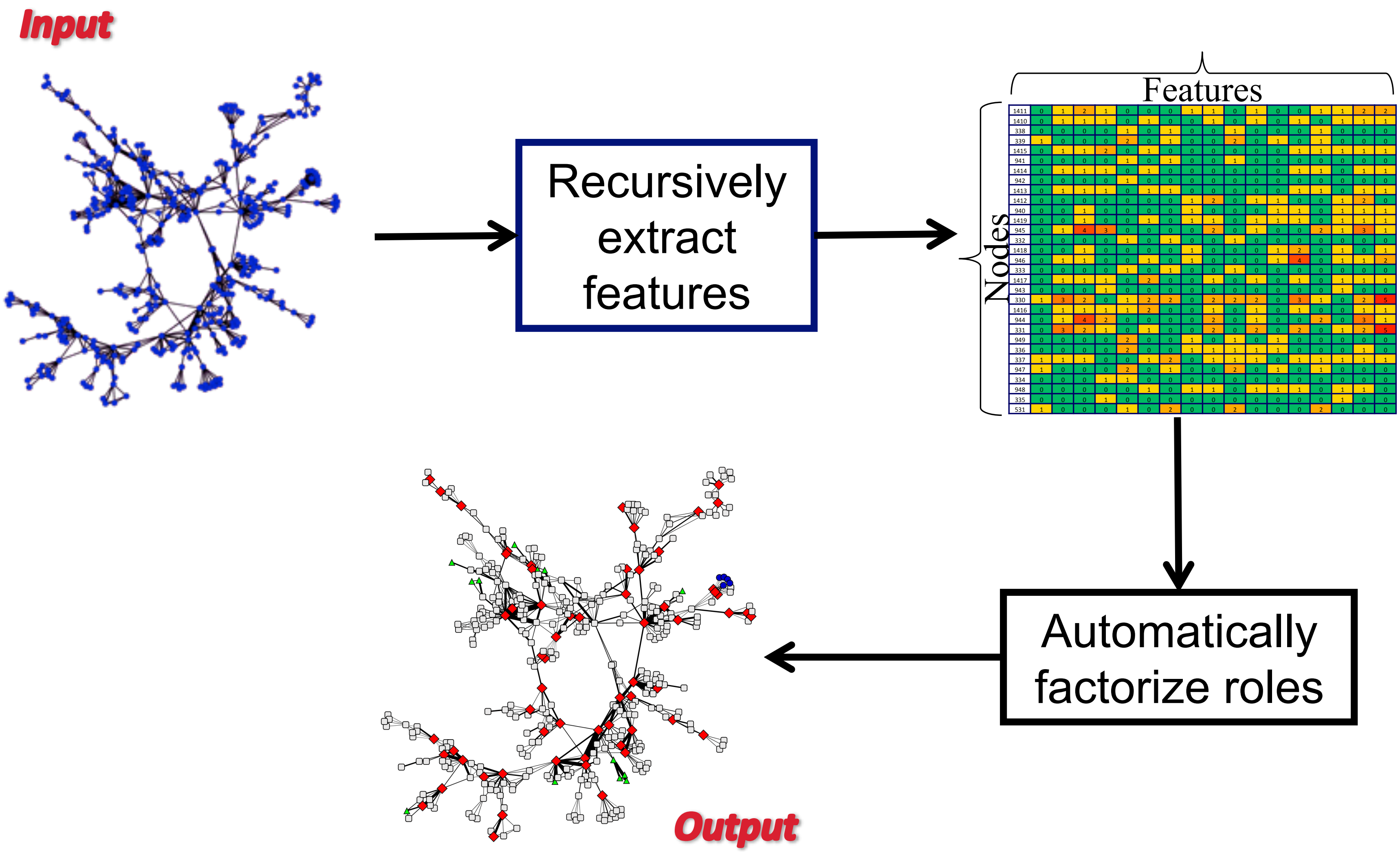
ReFeX (continued)

- Number of possible recursive features is infinite
- ReFeX pruning
  - Feature values are mapped to small integers via vertical logarithmic binning
    - Log binning places most of the discriminatory power among sets of nodes with large feature values



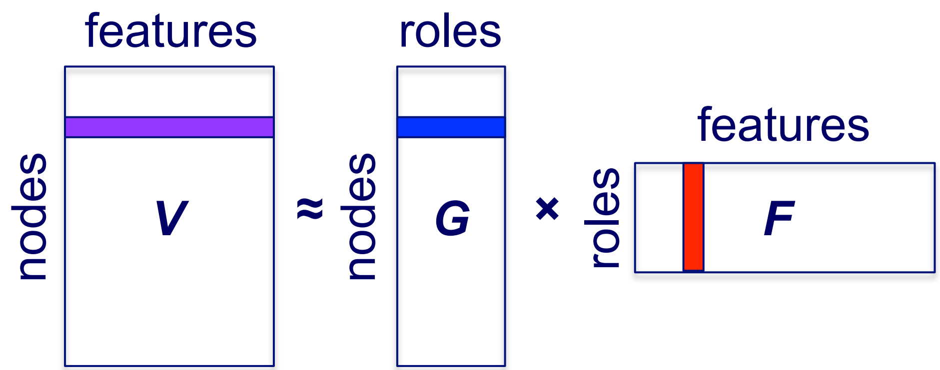
  - Look for pairs of features whose values never disagree by more than a threshold
  - A graph based approach
  - Threshold automatically set
  - Details in the KDD'11 paper

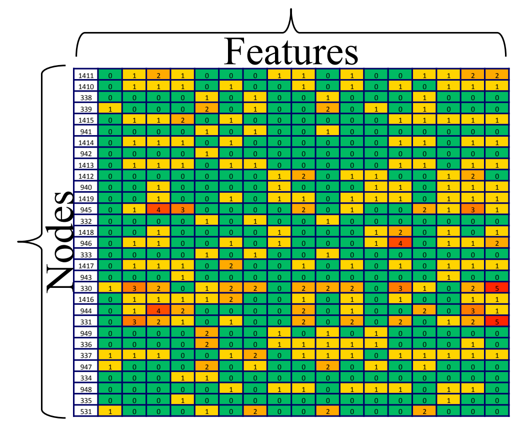
Role Extraction



Role Extraction: Feature Grouping

- Soft clustering in the structural feature space
  - Each node has a mixed-membership across roles
- Generate a rank *r* approximation of V ≈ GF





- RolX uses NMF for feature grouping
  - Computationally efficient
  - Non-negative factors simplify interpretation of roles and memberships

$$\mathrm{argmin}_{G,F} \|V - GF\|_{fro}, \text{s.t. } G \geq 0,\ F \geq 0$$

Role Extraction: Model Selection

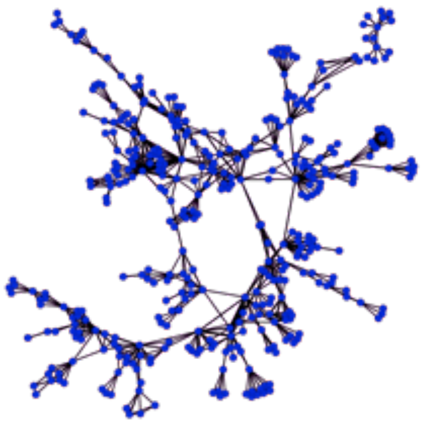
- Roles summarize behavior
  - Or, they compress the feature matrix, V
- Use MDL to select the model size *r* that results in the best compression
  - L: description length
  - M: # of bits required to describe the model
  - E: cost of describing the reconstruction errors in V – GF
  - Minimize L = M + E
    - To compress high-precision floating point values, RolX combines Llyod-Max quantization with Huffman codes
    - Errors in V-GF are not distributed normally, RolX uses KL divergence to compute E

$$\boxed{M = \bar{b}r(n+f)}$$

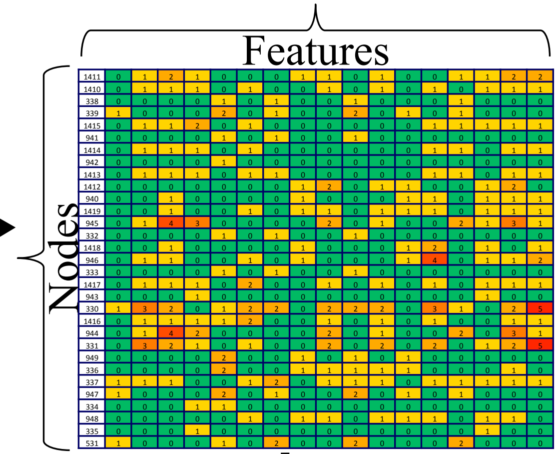
$$\boxed{E = \sum_{i,j}\left(V_{i,j}\log\frac{V_{i,j}}{(GF)_{i,j}} - V_{i,j} + (GF)_{i,j}\right)}$$

Role Extraction

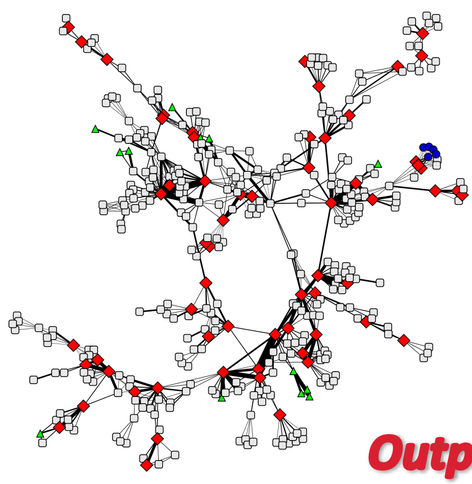
Input



Recursively extract features



Automatically factorize roles



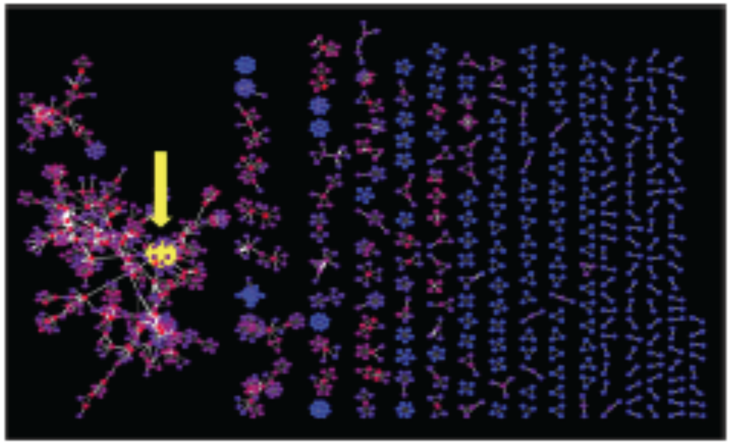
Output

Experiments on Role Discovery

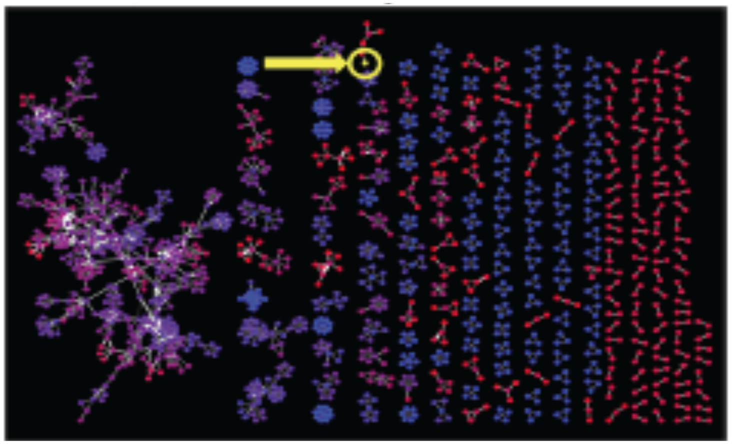
- Role query
- Role transfer
- Role sense-making
- Role mixed-memberships

Details in Henderson *et al.* KDD 2012

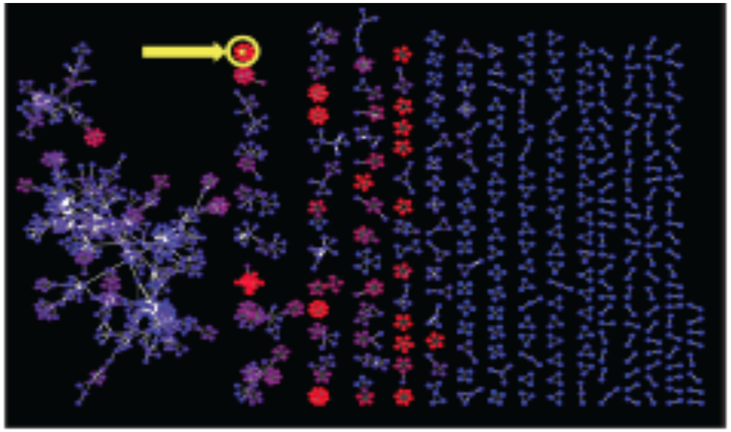
Role Query



Node Similarity for M.E.J. Newman (***bridge***)



Node Similarity for J. Rinzel (***isolated***)

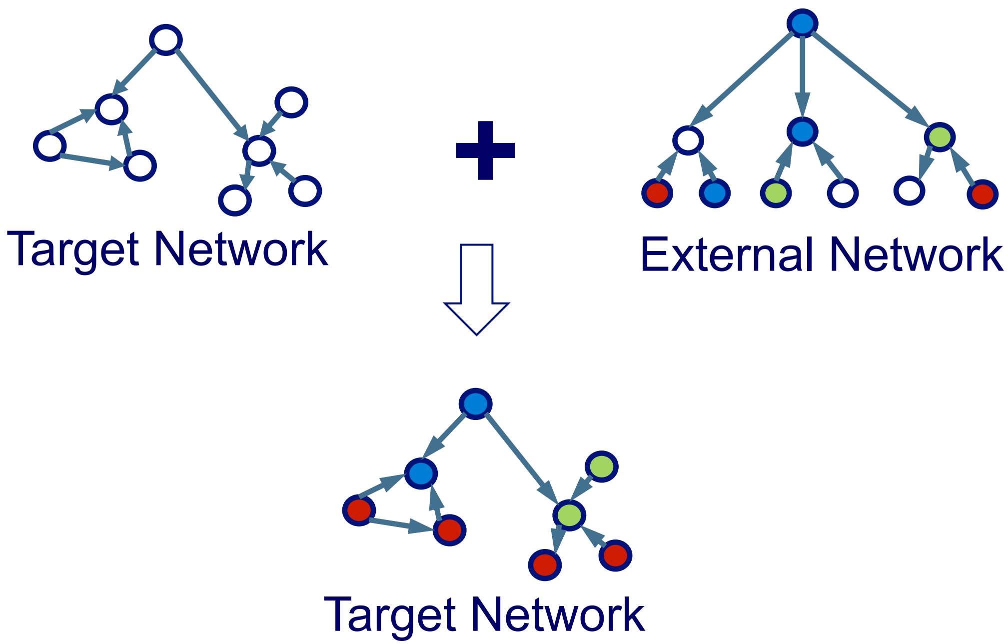


Node Similarity for F. Robert (***cliquey***)

Mixed-membership roles enable us to measure similarity of nodes based on their role memberships

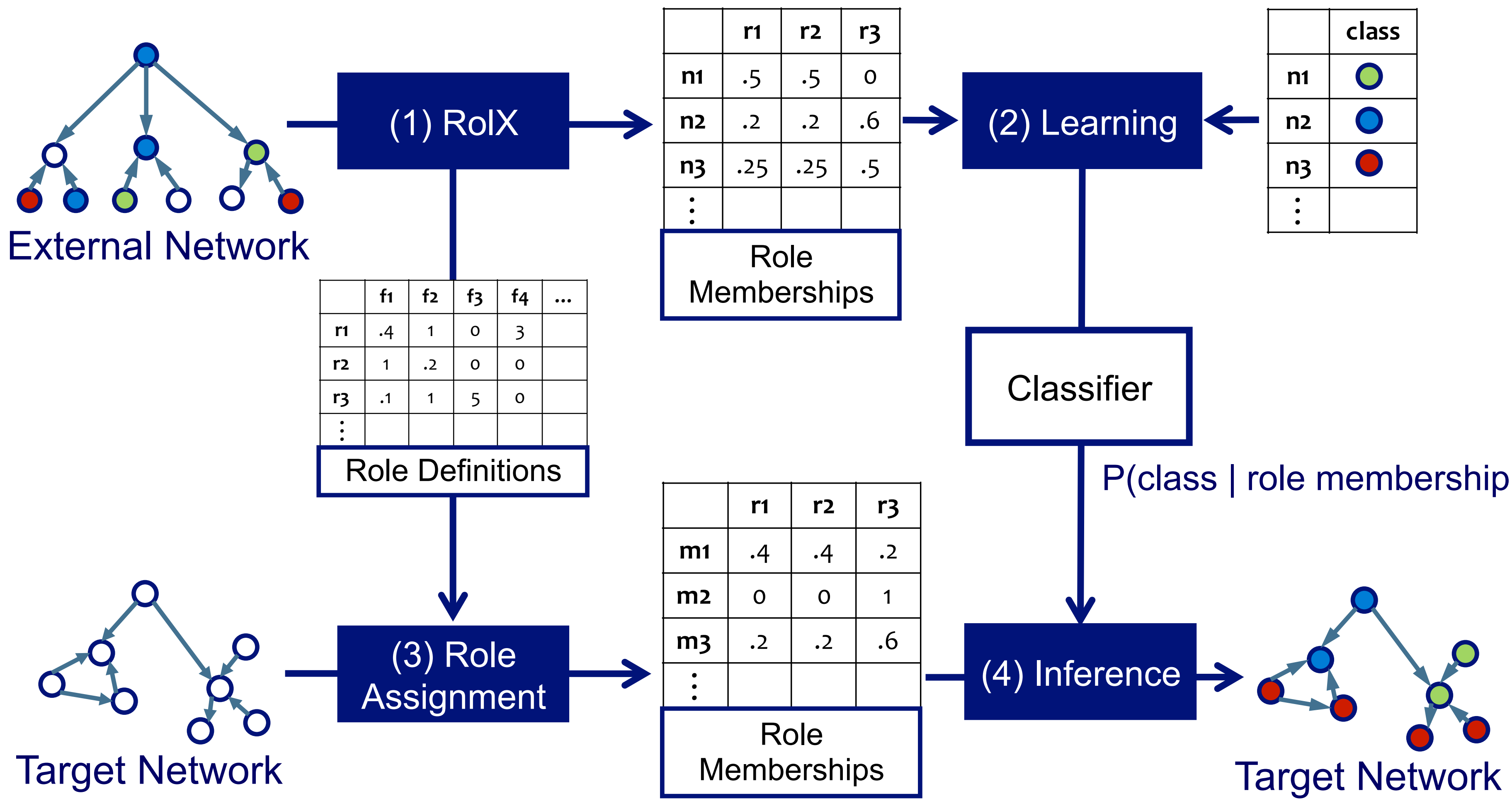
Role Transfer

- Question: How can we use labels from an external source to predict labels on a network with no labels?



- Conjecture: Nodes with similar roles are likely to have similar labels

Role Transfer = RolX + SL



Role Memberships

	r1	r2	r3
n1	.5	.5	0
n2	.2	.2	.6
n3	.25	.25	.5
⋮			

	class
n1	
n2	
n3	
⋮	

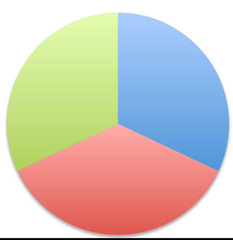
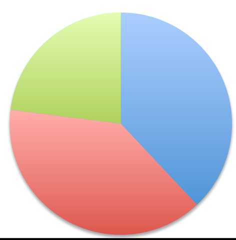
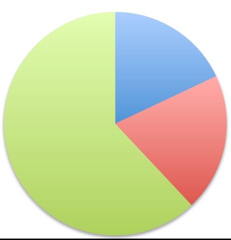
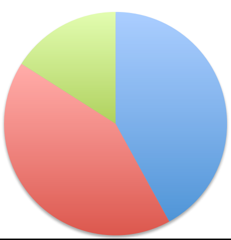
Role Definitions

	f1	f2	f3	f4	…
r1	.4	1	0	3	
r2	1	.2	0	0	
r3	.1	1	5	0	
⋮					

Role Memberships

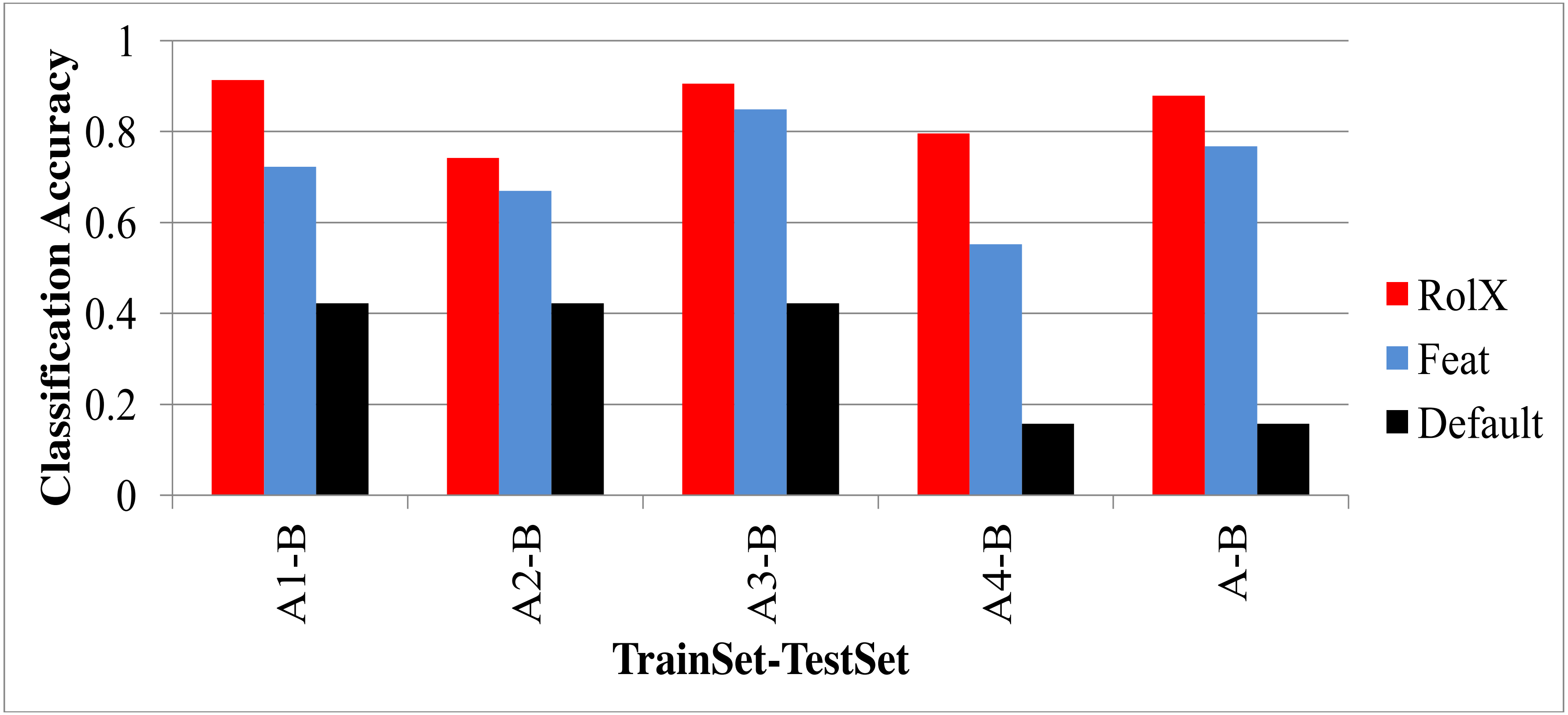
	r1	r2	r3
m1	.4	.4	.2
m2	0	0	1
m3	.2	.2	.6
⋮			

Data for Role Transfer

	IP-A1	IP-A2	IP-A3	IP-A4	IP-B
# Nodes	81,450	57,415	154,103	206,704	181,267
% labeled	36.7%	28.1%	20.1%	32.9%	15.3%
# Links	968,138	432,797	1,266,341	1,756,082	1,945,215
(# unique)	206,112	137,822	358,851	465,869	397,925
Class Distribu-tion					

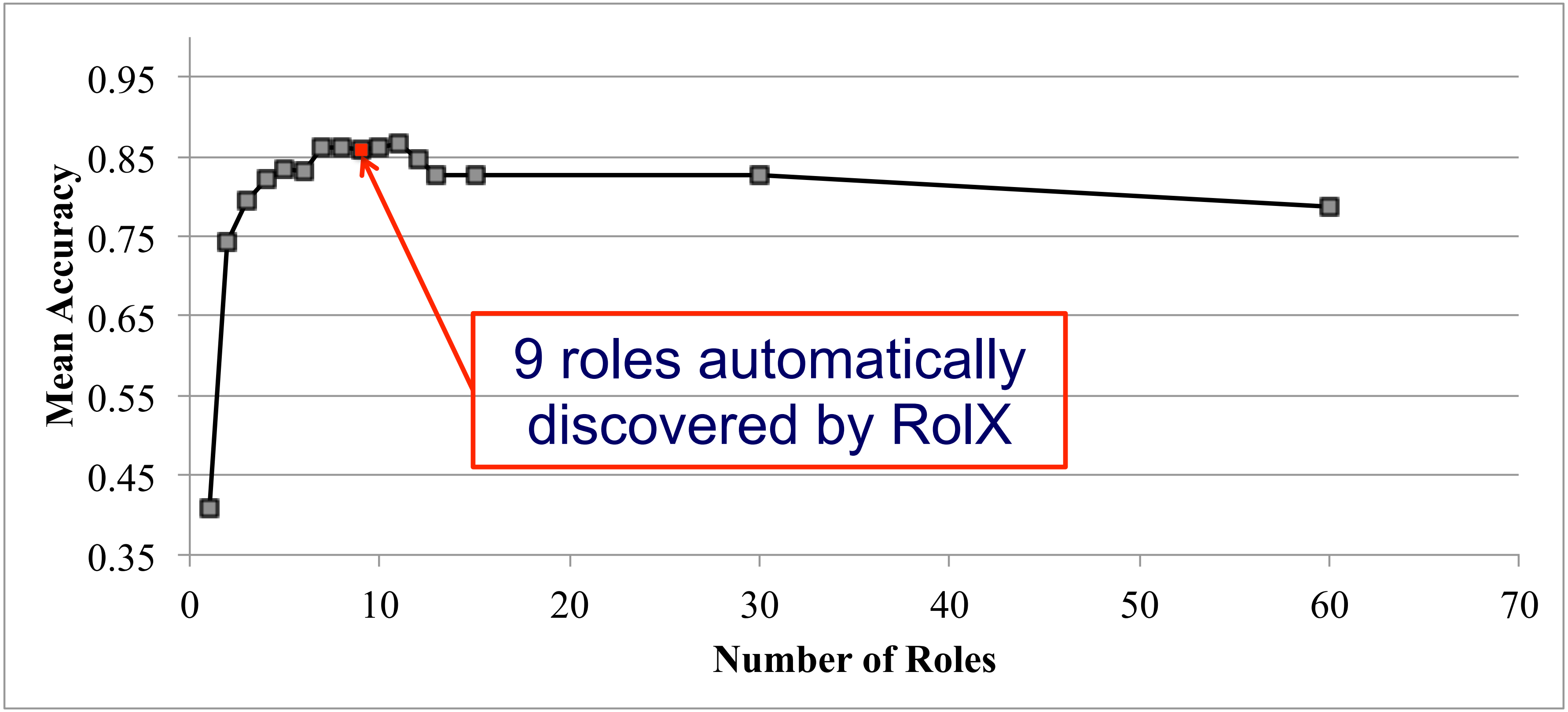
Web DNS P2P

Role Transfer Results



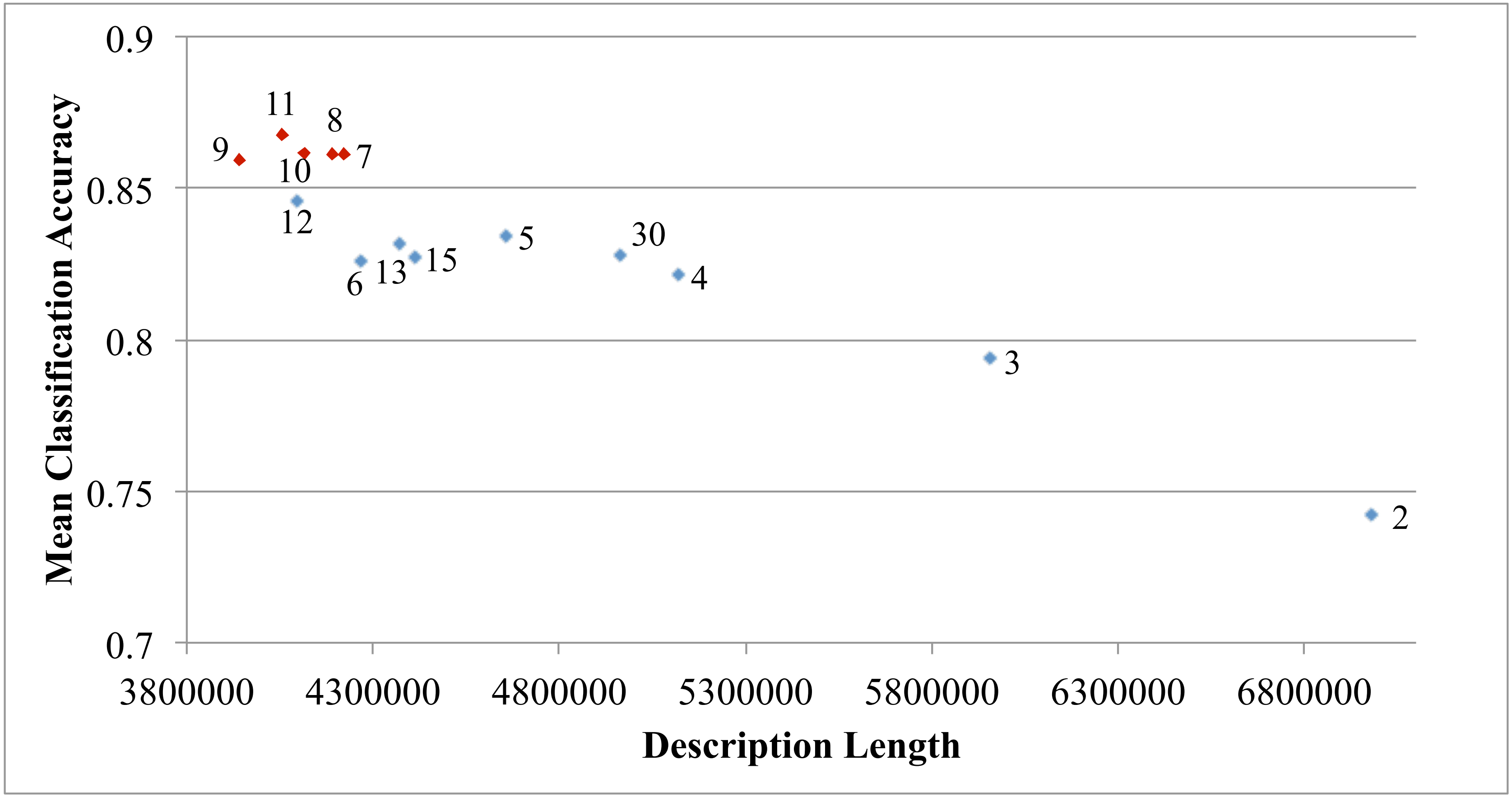
Roles generalize across disjoint networks & enable prediction without re-learning

Model Selection



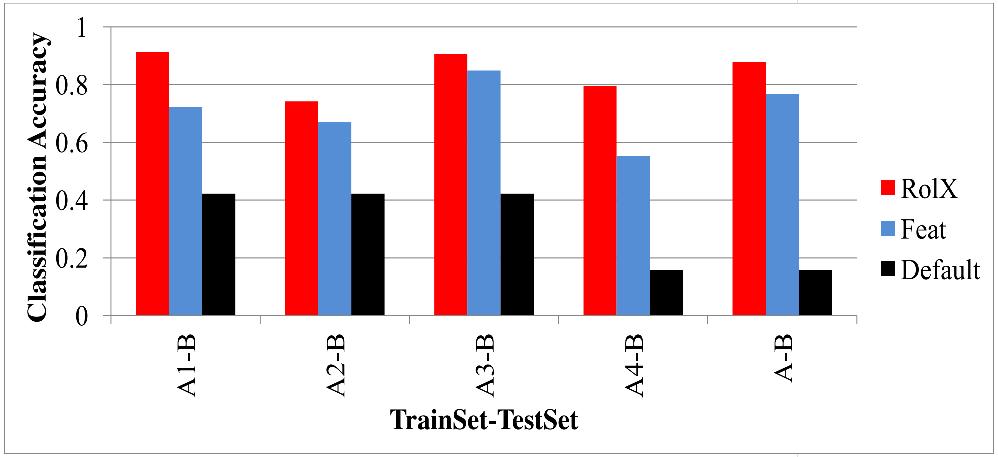
RolX selects high accuracy model sizes

Model Selection (continued)

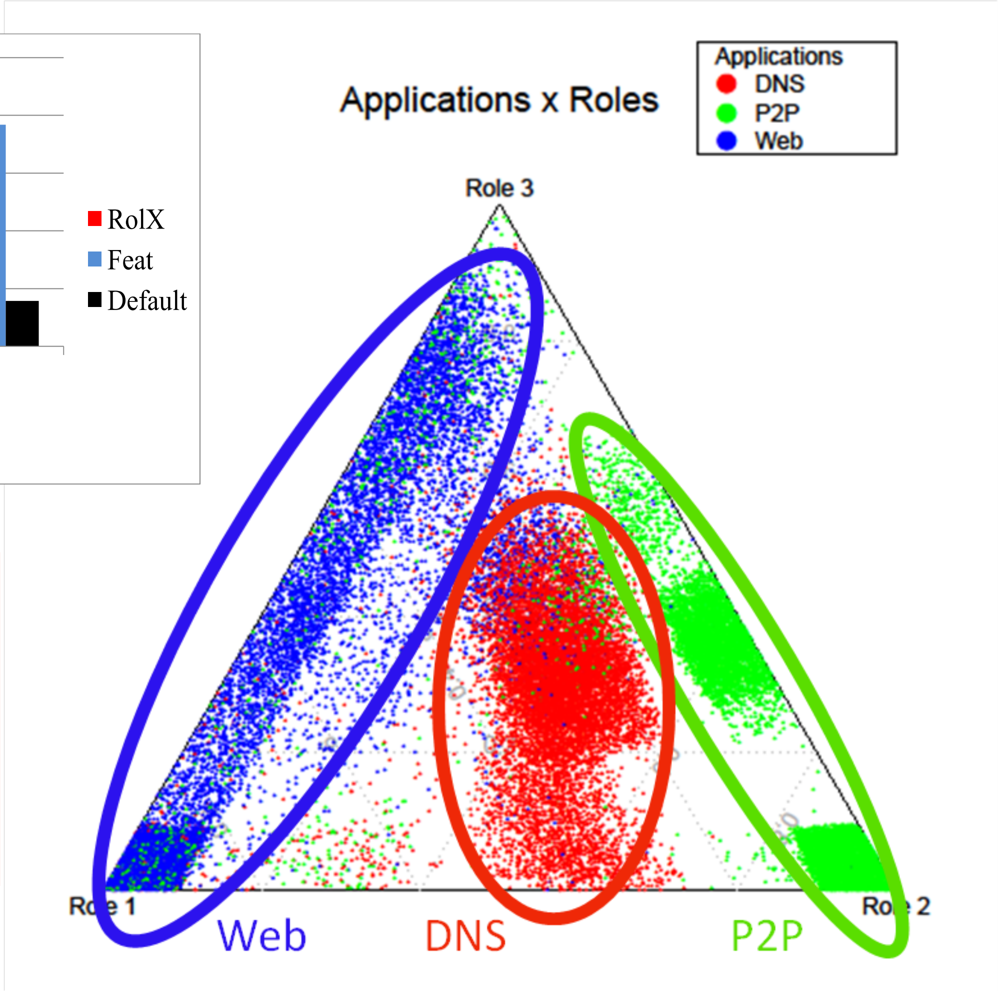


Classification accuracy is highest when RolX selection criterion is minimized

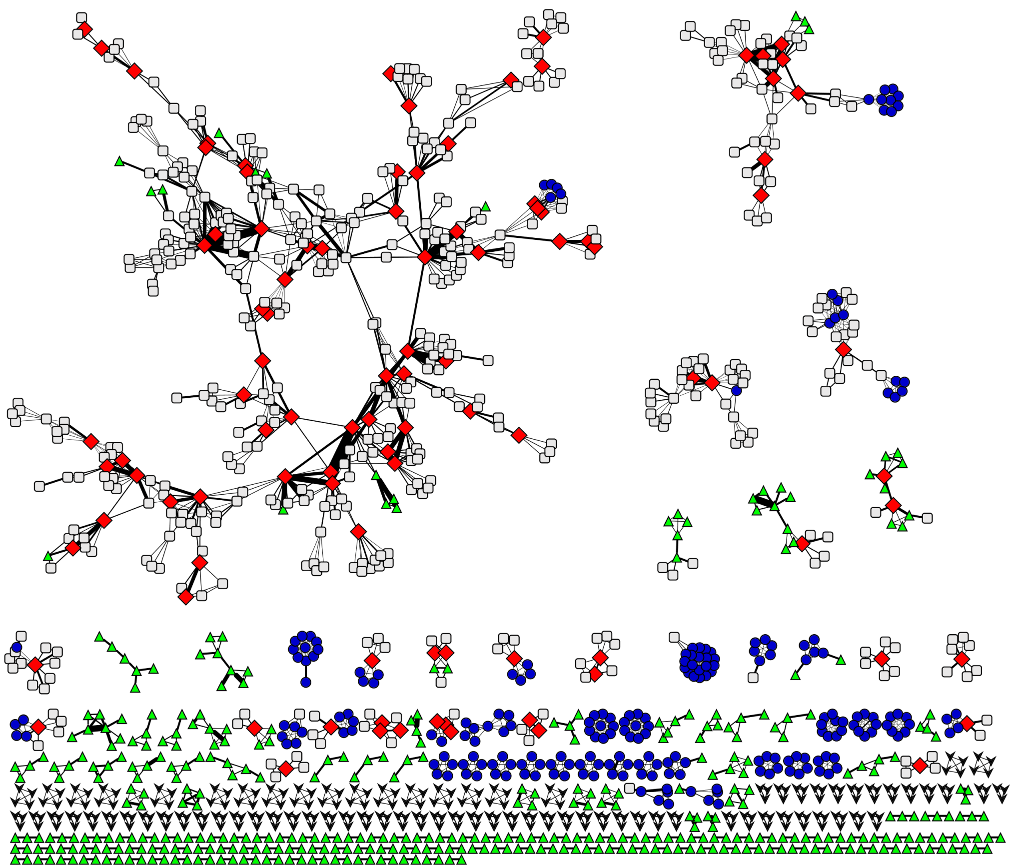
Role Space



IP trace classes are well-separated in the RolX **role space** with as few as 3 roles

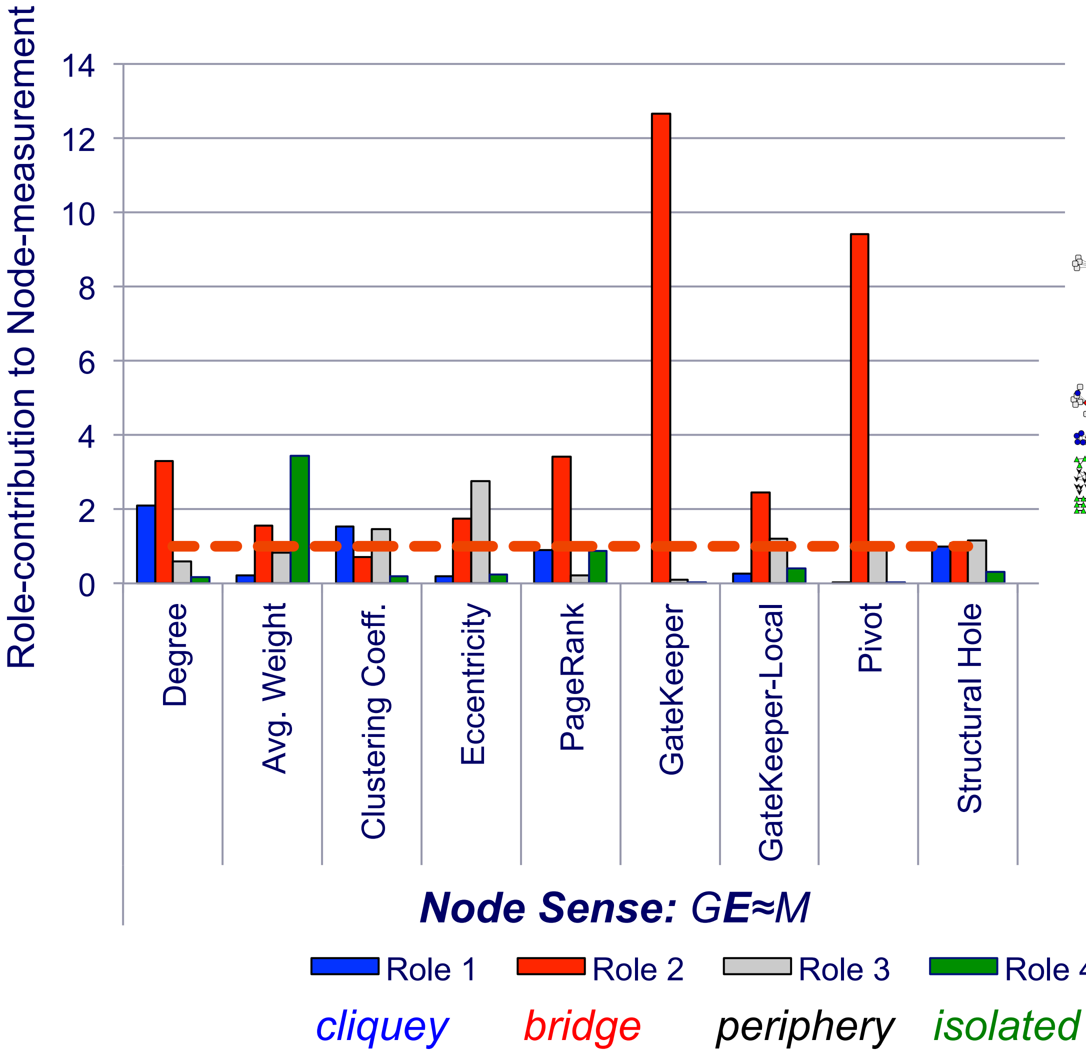


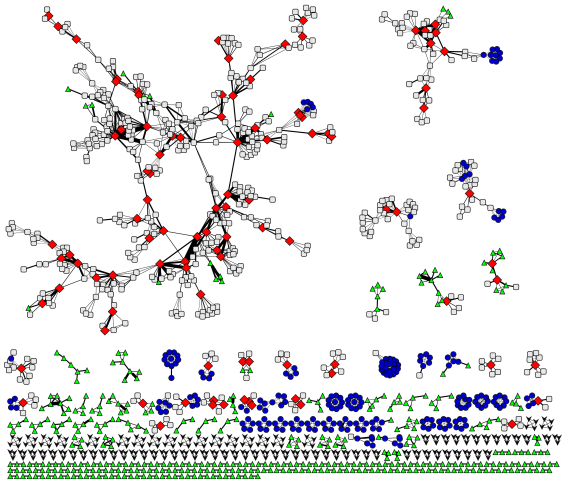
Automatically Discovered Roles



Network Science Co-authorship Graph
[Newman 2006]

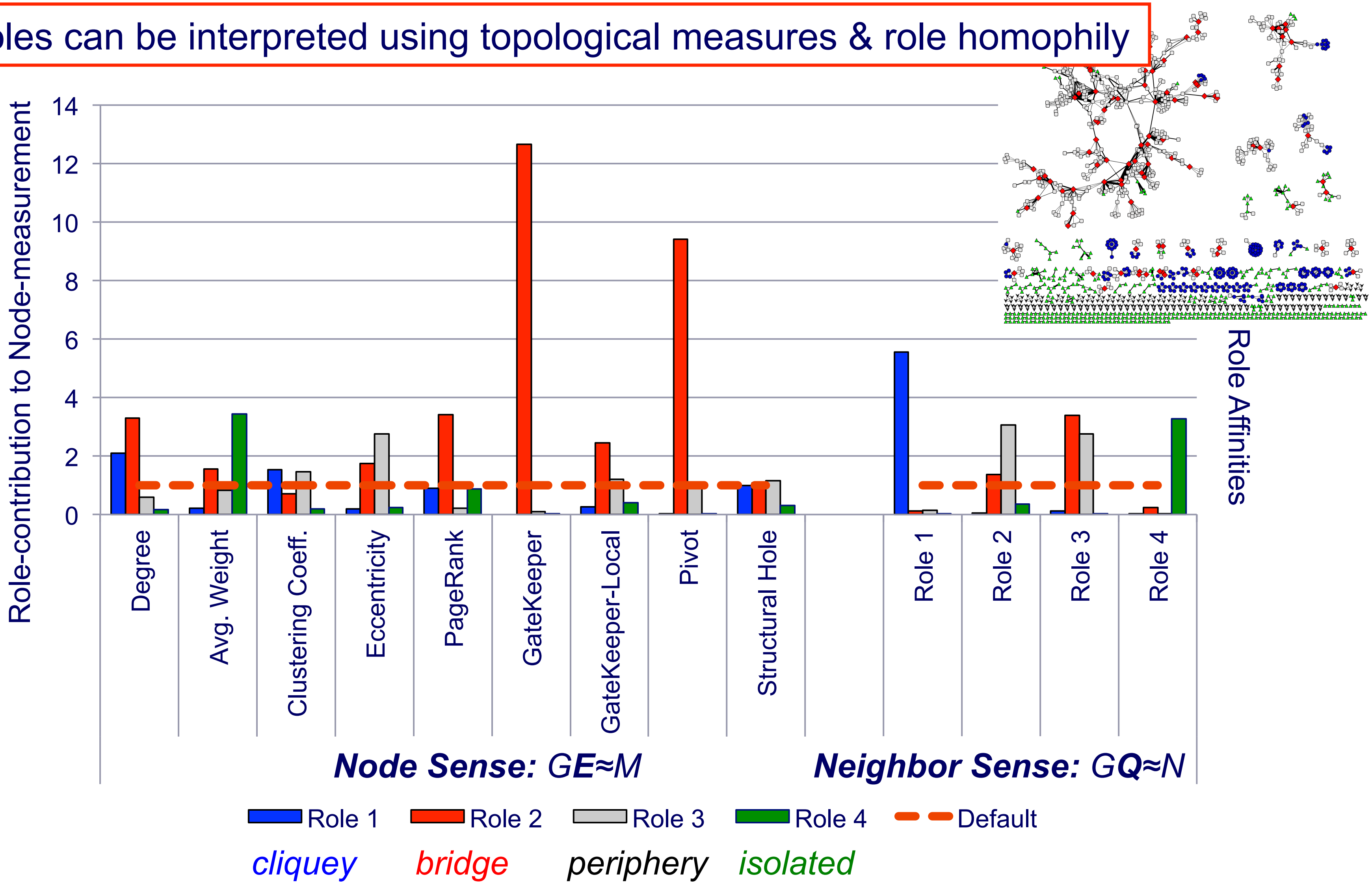
Making Sense of Roles





Making Sense of Roles

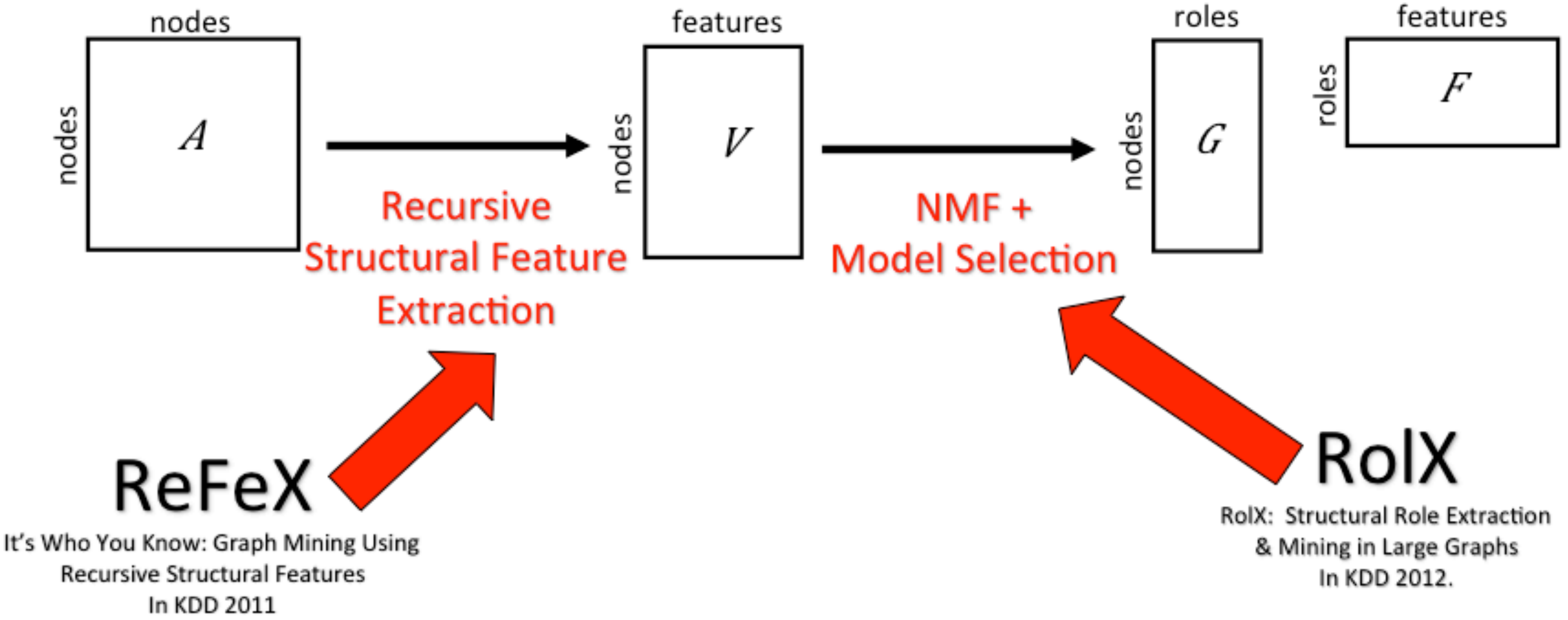
Roles can be interpreted using topological measures & role homophily



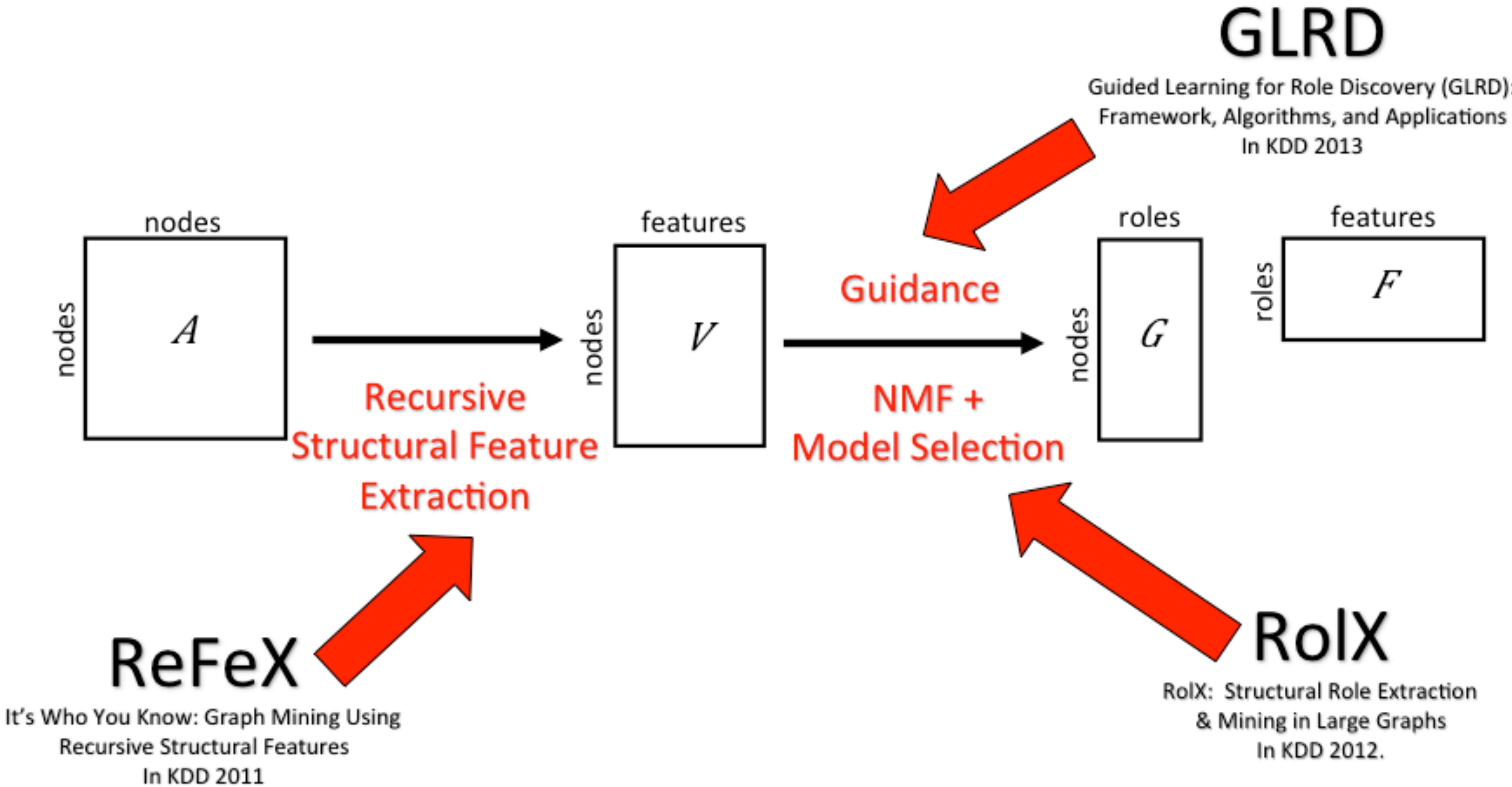
GLRD: Guided Learning for Role Discovery

- Introduced by Sean Gilpin et al.
- RolX is unsupervised
- What if we had guidance on roles?
  - Guidance as in weak supervision encoded as constraints
- Types of guidance
  - Sparse roles
  - Diverse roles
  - Alternative roles, given a set of existing roles

GLRD


nodes
nodes
A
Recursive
Structural Feature
Extraction
features
nodes
V
NMF +
Model Selection
roles
nodes
G
features
roles
F
ReFeX
It's Who You Know: Graph Mining Using
Recursive Structural Features
In KDD 2011
RolX
RolX: Structural Role Extraction
& Mining in Large Graphs
In KDD 2012.

GLRD



GLRD
Guided Learning for Role Discovery (GLRD):
Framework, Algorithms, and Applications
In KDD 2013
nodes
nodes
A
Recursive
Structural Feature
Extraction
features
nodes
V
Guidance
NMF +
Model Selection
roles
nodes
G
features
roles
F
ReFeX
It's Who You Know: Graph Mining Using
Recursive Structural Features
In KDD 2011
RolX
RolX: Structural Role Extraction
& Mining in Large Graphs
In KDD 2012.

GLRD Framework

- Constraints on columns of G (i.e., role assignments) or rows of F (i.e. role definitions) are convex functions

$$
\begin{aligned}
\underset{\mathbf{G},\mathbf{F}}{\text{minimize}} \quad & ||\mathbf{V} - \mathbf{G}\mathbf{F}||_2 \\
\text{subject to} \quad & g_i(\mathbf{G}) \leq d_{Gi}, \; i = 1, \ldots, t_G \\
& f_i(\mathbf{F}) \leq d_{Fi}, \; i = 1, \ldots, t_F
\end{aligned}
$$

where g_i and f_i are convex functions.

- Use an alternative least squares (ALS) formulation
  - Do not alternate between solving for the entire G and F
  - Solve for one column of G or one row of F at a time
    - This is okay since we have convex constraints

Guidance Overview

Guidance Type	Effect of increasing guidance	
	on role assignment (*G*)	on role definition (*F*)
Sparsity	Reduces the number of nodes with minority memberships in roles	Decreases likelihood that features with small explanatory benefit are included
Diversity	Limits the amount of allowable overlap in assignments	Roles must be explained with completely different sets of features
Alternative	Decreases the allowable similarity between the two sets of role assignments	Ensures that role definitions are very dissimilar between the two sets of role assignments

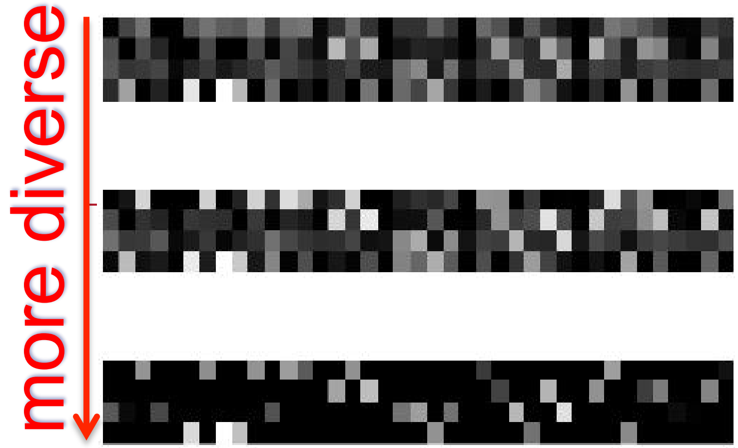
Sparsity

$$
\begin{aligned}
\underset{\mathbf{G},\mathbf{F}}{\text{argmin}} \quad & ||\mathbf{V} - \mathbf{GF}||_2 \\
\text{subject to:} \quad & \mathbf{G} \geq 0, \mathbf{F} \geq 0 \\
& \forall i \quad ||\mathbf{G_{\bullet i}}||_1 \leq \epsilon_G \\
& \forall i \quad ||\mathbf{F_{i \bullet}}||_1 \leq \epsilon_F
\end{aligned}
$$

where ϵ_G and ϵ_F define upperbounds for the sparsity constraints (amount of allowable density).

Diversity

Goal: Find role assignments or definitions that are very different from each other



$$\underset{\mathbf{G},\mathbf{F}}{\text{argmin}} \quad ||\mathbf{V} - \mathbf{GF}||_2$$

$$\text{subject to:} \quad \mathbf{G} \geq 0, \mathbf{F} \geq 0$$

$$\forall i, j \quad \mathbf{G}_{\bullet i}^T \mathbf{G}_{\bullet j} \leq \epsilon_G \quad i \neq j$$

$$\forall i, j \quad \mathbf{F}_{i\bullet}\, \mathbf{F}_{j\bullet}^T \leq \epsilon_F \quad i \neq j$$

where ϵ_G and ϵ_F define upperbounds on how angularly similar role assignments and role definitions can be to each other.

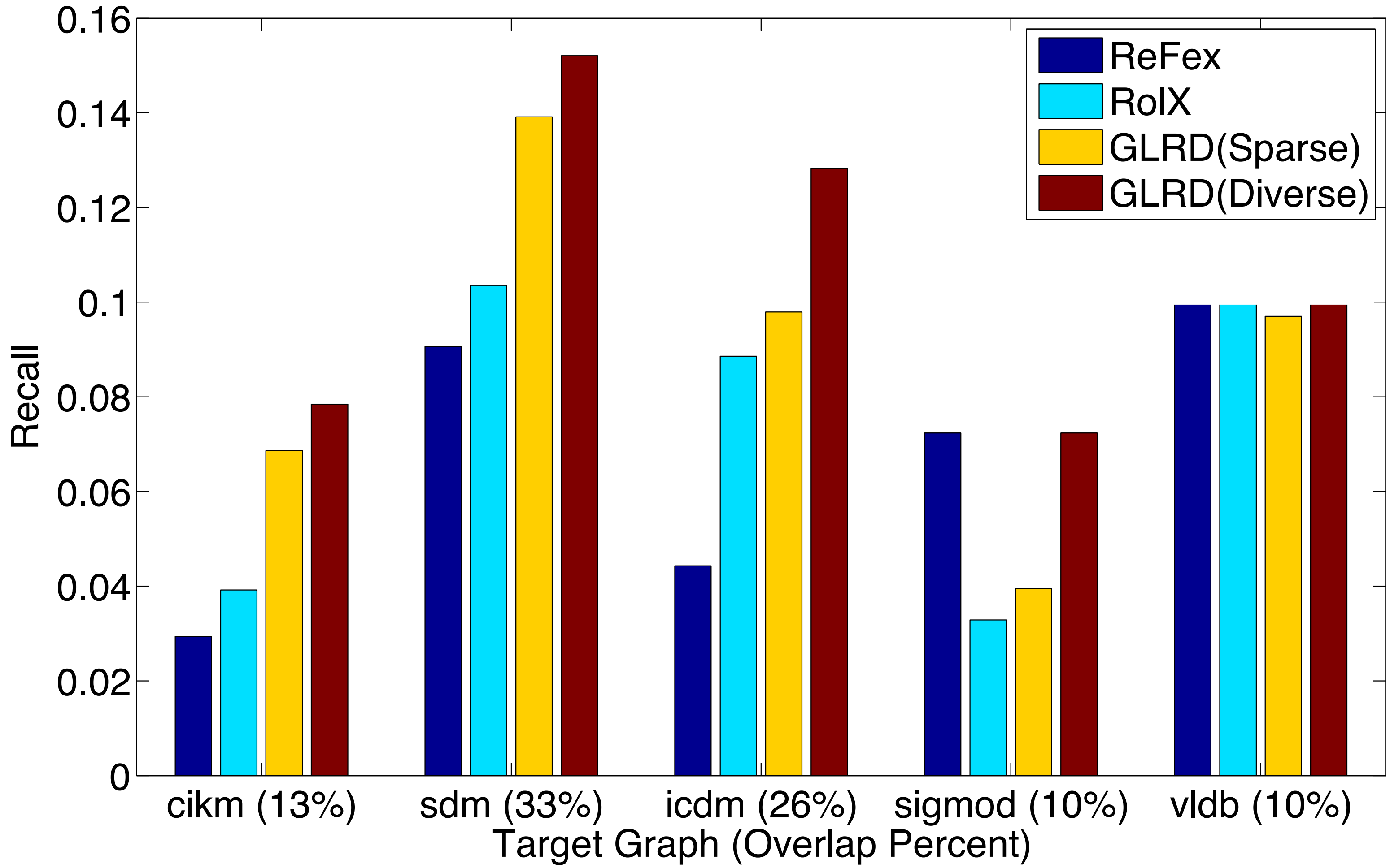
Diverse Roles and Sparse Roles

- Question: Can diversity and sparsity constraints create better role definitions?
- Conjecture: Better role definitions will better facilitate other problems such as identity resolution across graphs
- Experiment: Compare graph mining results using various methods for role discovery

Network	$\|V\|$	$\|E\|$	k	$\|LCC\|$	#CC
VLDB	1,306	3,224	4.94	769	112
SIGMOD	1,545	4,191	5.43	1,092	116
CIKM	2,367	4,388	3.71	890	361
SIGKDD	1,529	3,158	4.13	743	189
ICDM	1,651	2,883	3.49	458	281
SDM	915	1,501	3.28	243	165

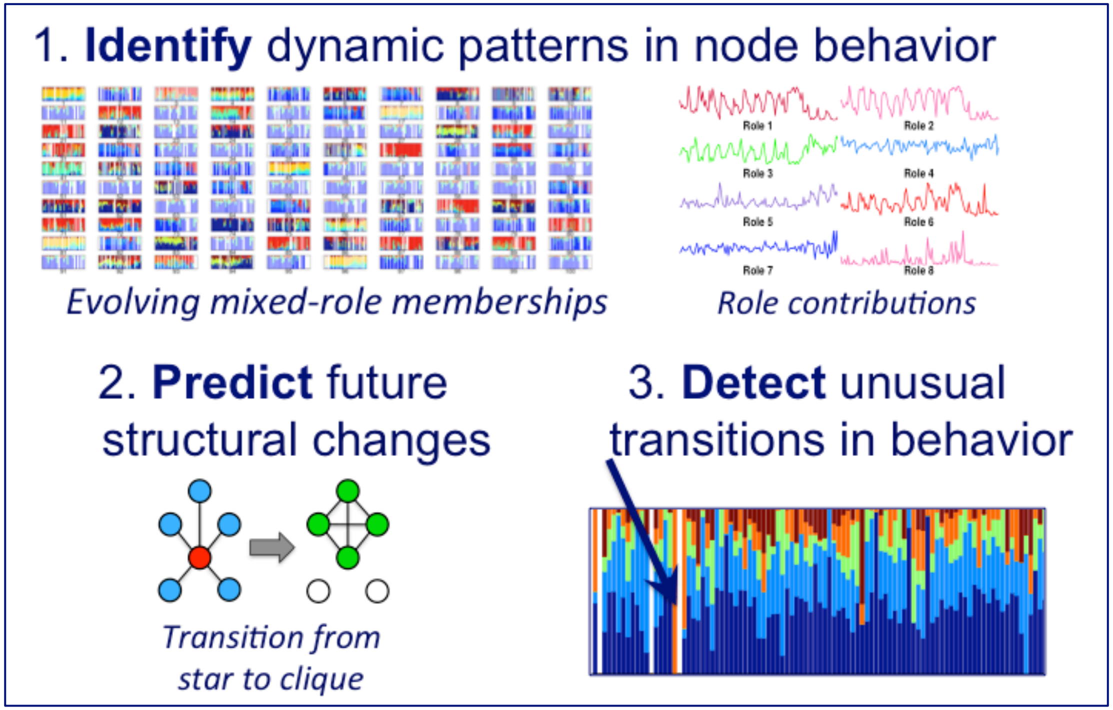
DBLP Co-authorship Networks from 2005-2009

Identity Resolution across Networks



Modeling Dynamic Graphs with Roles

- Introduced by Rossi et al. [WSDM 2013]



- Given G_{t-1} and G_t find a transition model T that minimizes the functional:

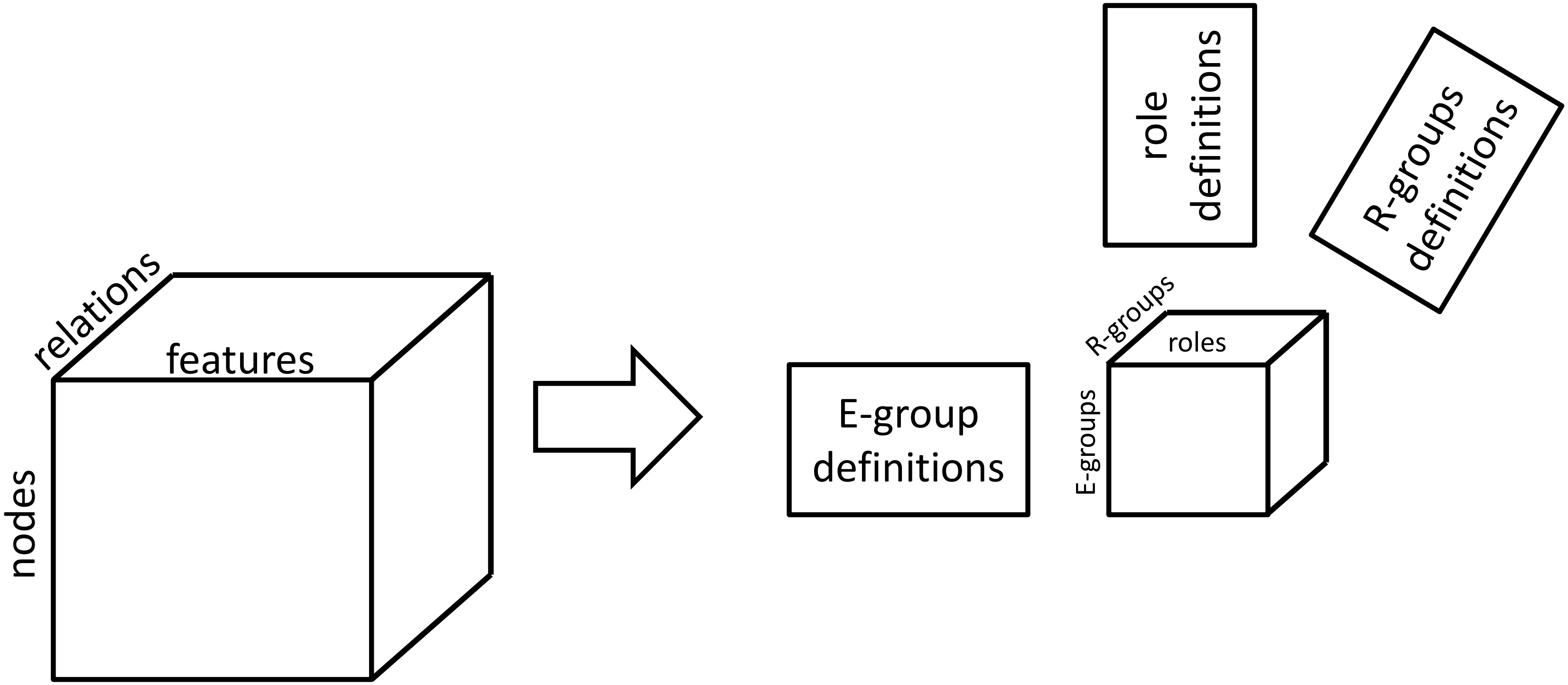
$$f(\mathbf{G}_t, \mathbf{G}_{t-1}) = \frac{1}{2}||\mathbf{G}_t - \mathbf{G}_{t-1}\mathbf{T}||_F^2$$

- All models predict G_{t+1} using G_t as

$$\mathbf{G'_{t+1}} = \mathbf{G_t T}$$

Roles Across Relations

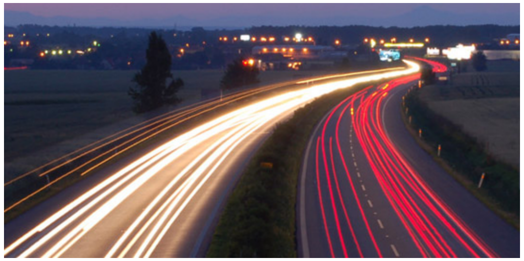
- Role Discovery in Multi-Relational Graphs [Sean Gilpin, et al. in preparation]



Two New Hybrid Approaches

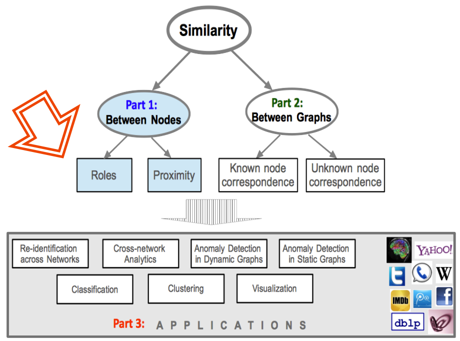
- Ryan Rossi and Nesreen K. Ahmed
  - Role Discovery in Networks [*TKDE* 2014]
  - A taxonomy for discovering roles that includes (*i*) graph-based roles, (*ii*) feature-based roles, and (*iii*) hybrid roles
- Yiye Ruan and Srinivasan Parthasarathy
  - Simultaneous Detection of Communities and Roles from Large Networks [*COSN* 2014]
  - *RC-Joint*: a non-parametric approach to simultaneously identify softly assigned communities and structural roles

Roadmap



- Node Roles
  - What are roles
  - Roles and communities
  - Roles and equivalences (from sociology)
  - Roles (from data mining)
  - Summary
- Node Proximity (after the coffee break)

Summary of Part 1a on Roles



- Are structural behavior ("function") of nodes
- Are complementary to communities
- Previous work mostly in sociology under equivalences
- Recent graph mining work produces mixed-membership roles, is fully automatic and scalable
- Can be used for many tasks: transfer learning, re-identification, anomaly detection, etc
- Extensions: including guidance, modeling dynamic networks, etc

Role Discovery vs. Regular Equivalence

	Role Discovery	Regular Equivalence
Mixed-membership over roles	✓	
Automatically selects the best model	✓	
Can incorporate arbitrary features	✓	
Uses structural features	✓	
Uses structure	✓	✓
Generalizes across disjoint networks (longitudinal & cross-sectional)	✓	?
Scalable (linear on # of edges)	✓	
Guidance	✓	

Acknowledgement

- LLNL: Brian Gallagher, Keith Henderson
- ASU: Hanghang Tong
- Google: Sugato Basu
- SUNY Stony Brook: Leman Akoglu
- CMU: Christos Faloutsos, Danai Koutra
- UC Berkeley: Lei Li
- UC Davis: Ian Davidson, Sean Gilpin
- Rutgers: Long Le

Thanks to: LLNL, NSF, IARPA, DARPA, DTRA.

Papers at http://eliassi.org/pubs.html

- Long T. Le, Tina Eliassi-Rad, Hanghang Tong: Learning to minimize dissemination on large graphs. **under review**, 2014.
- Sean Gilpin, Tom Kuo, Tina Eliassi-Rad, Ian Davidson: Roles across relations: Role discovery in multi-relational graphs. **under review**, 2014.
- Michele Berlingerio, Danai Koutra, Tina Eliassi-Rad, Christos Faloutsos: Network similarity via multiple social theories. **ASONAM** 2013: 1439-1440.
- Sean Gilpin, Tina Eliassi-Rad, Ian Davidson: Guided learning for role discovery (GLRD): Framework, algorithms, and applications. **KDD** 2013: 113-121.
- Ryan A. Rossi, Brian Gallagher, Jennifer Neville, Keith Henderson: Modeling dynamic behavior in large evolving graphs. **WSDM** 2013: 667-676. http://www.ryanrossi.com/papers/wsdm13-dbmm.pdf
- Hanghang Tong, B. Aditya Prakash, Tina Eliassi-Rad, Michalis Faloutsos, Christos Faloutsos: Gelling, and melting, large graphs by edge manipulation. **CIKM** 2012: 245-254.
- Keith Henderson, Brian Gallagher, Tina Eliassi-Rad, Hanghang Tong, Sugato Basu, Leman Akoglu, Danai Koutra, Christos Faloutsos, Lei Li: RolX: Structural role extraction & mining in large graphs. **KDD** 2012: 1231-1239.
- Ryan A. Rossi, Brian Gallagher, Jennifer Neville, Keith Henderson: Role-dynamics: fast mining of large dynamic networks. **WWW** (Companion Volume) 2012: 997-1006.
- Keith Henderson, Brian Gallagher, Lei Li, Leman Akoglu, Tina Eliassi-Rad, Hanghang Tong, Christos Faloutsos: It's who you know: Graph mining using recursive structural features. **KDD** 2011: 663-671.

References

Deterministic Equivalences

- S. Boorman, H.C. White: Social Structure from Multiple Networks: II. Role Structures. American Journal of Sociology, 81:1384-1446, 1976.
- S.P. Borgatti, M.G. Everett: Notions of Positions in Social Network Analysis. In P. V. Marsden (Ed.): Sociological Methodology, 1992:1–35.
- S.P. Borgatti, M.G. Everett, L. Freeman: UCINET IV, 1992.
- S.P. Borgatti, M.G. Everett, Regular Blockmodels of Multiway, Multimode Matrices. Social Networks, 14:91-120, 1992.
- R. Breiger, S. Boorman, P. Arabie: An Algorithm for Clustering Relational Data with Applications to Social Network Analysis and Comparison with Multidimensional Scaling. Journal of Mathematical Psychology, 12:328-383, 1975.
- R.S. Burt: Positions in Networks. Social Forces, 55:93-122, 1976.

References

- P. DiMaggio: Structural Analysis of Organizational Fields: A Blockmodel Approach. Research in Organizational Behavior, 8:335-70, 1986.
- P. Doreian, V. Batagelj, A. Ferligoj: Generalized Blockmodeling. Cambridge University Press, 2005.
- M.G. Everett, S. P. Borgatti: Regular Equivalence: General Theory. Journal of Mathematical Sociology, 19(1):29-52, 1994.
- K. Faust, A.K. Romney: Does Structure Find Structure? A critique of Burt's Use of Distance as a Measure of Structural Equivalence. Social Networks, 7:77-103, 1985.
- K. Faust, S. Wasserman: Blockmodels: Interpretation and Evaluation. Social Networks, 14:5–61. 1992.
- R.A. Hanneman, M. Riddle: Introduction to Social Network Methods. University of California, Riverside, 2005.

References

- F. Lorrain, H.C. White: Structural Equivalence of Individuals in Social Networks. Journal of Mathematical Sociology, 1:49-80, 1971.
- L.D. Sailer: Structural Equivalence: Meaning and Definition, Computation, and Applications. Social Networks, 1:73-90, 1978.
- M.K. Sparrow: A Linear Algorithm for Computing Automorphic Equivalence Classes: The Numerical Signatures Approach. Social Networks, 15:151-170, 1993.
- S. Wasserman, K. Faust: Social Network Analysis: Methods and Applications. Cambridge University Press, 1994.
- H.C. White, S. A. Boorman, R. L. Breiger: Social Structure from Multiple Networks I. Blockmodels of Roles and Positions. American Journal of Sociology, 81:730-780, 1976.
- D.R. White, K. Reitz: Graph and Semi-Group Homomorphism on Networks and Relations. Social Networks, 5:143-234, 1983.

References

Stochastic Blockmodels

- E.M. Airoldi, D.M. Blei, S.E. Fienberg, E.P. Xing: Mixed Membership Stochastic Blockmodels. Journal of Machine Learning Research, 9:1981-2014, 2008.
- P.W. Holland, K.B. Laskey, S. Leinhardt: Stochastic Blockmodels: Some First Steps. Social Networks, 5:109-137, 1983.
- C. Kemp, J.B. Tenenbaum, T.L. Griffiths, T. Yamada, N. Ueda: Learning Systems of Concepts with an Infinite Relational Model. AAAI 2006.
- P.S. Koutsourelakis, T. Eliassi-Rad: Finding Mixed-Memberships in Social Networks. AAAI Spring Symposium on Social Information Processing, Stanford, CA, 2008.
- K. Nowicki ,T. Snijders: Estimation and Prediction for Stochastic Blockstructures, Journal of the American Statistical Association, 96:1077–1087, 2001.
- Z. Xu, V. Tresp, K. Yu, H.-P. Kriegel: Infinite Hidden Relational Models. UAI 2006.
- S. Wasserman, C. Anderson: Stochastic a Posteriori Blockmodels: Construction and Assessment, Social Networks, 9:1-36, 1987.

References

Role Discovery

- K. Henderson, B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, C. Faloutsos: It's Who Your Know: Graph Mining Using Recursive Structural Features. KDD 2011: 663-671.
- R. Jin, V. E. Lee, H. Hong: Axiomatic ranking of network role similarity. KDD 2011: 922-930.
- K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, C. Faloutsos, L. Li: RolX: Structural role extraction & mining in large graphs. KDD 2012: 1231-1239.
- R. A. Rossi, B. Gallagher, J. Neville, K. Henderson: Modeling dynamic behavior in large evolving graphs. WSDM 2013: 667-676.
- S. Gilpin, T. Eliassi-Rad, I. Davidson: Guided Learning for Role Discovery (GLRD): Framework, algorithms, and applications. KDD 2013.
- R.A. Rossi, N. K. Ahmed: Role Discovery in Networks. TKDE 2014.
- Y. Ruan, S. Parthasarathy: Simultaneous Detection of Communities and Roles from Large Networks. COSN 2014.

References

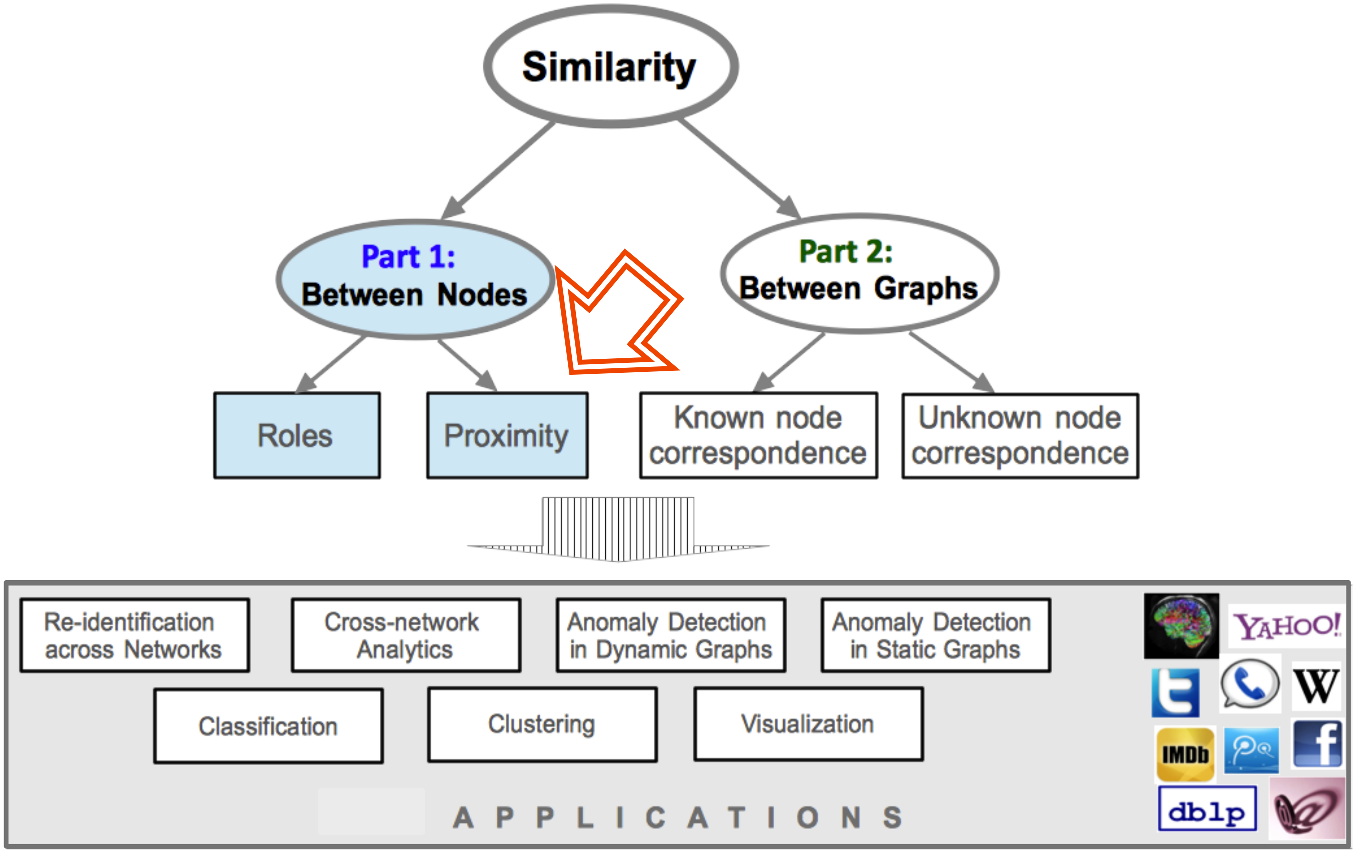
Community Discovery (only the ones mentioned in this tutorial)

- A. Clauset, M.E.J. Newman, C. Moore: Finding Community Structure in Very Large Networks. Phys. Rev. E., 70:066111, 2004.
- M.E.J. Newman: Finding Community Structure in Networks Using the Eigenvectors of Matrices. Phys. Rev. E., 74:036104, 2006.

Propositionalisation

- A.J. Knobbe, M. de Haas, A. Siebes: Propositionalisation and Aggregates. PKDD 2001: 277-288.
- M.-A. Krogel, S. Rawles, F. Zelezny, P.A. Flach, N. Lavrac, S. Wrobel: Comparative Evaluation of Approaches to Propositionalization. ILP 2003: 197-214.
- J. Neville, D. Jensen, B. Gallagher: Simple Estimators for Relational Bayesian Classifiers. ICDM 2003: 609-612.

Next



Similarity
Part 1:
Between Nodes
Part 2:
Between Graphs
Roles
Proximity
Known node correspondence
Unknown node correspondence
Re-identification across Networks
Cross-network Analytics
Anomaly Detection in Dynamic Graphs
Anomaly Detection in Static Graphs
Classification
Clustering
Visualization
APPLICATIONS
YAHOO!
W
IMDb
dblp