Discovering Roles and Anomalies in Graphs: Theory and Applications

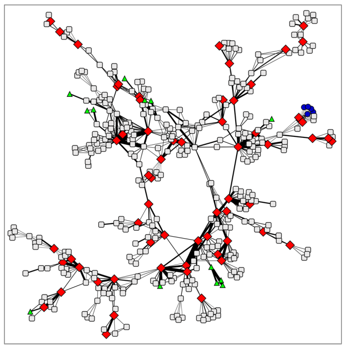
Part 1: Theory

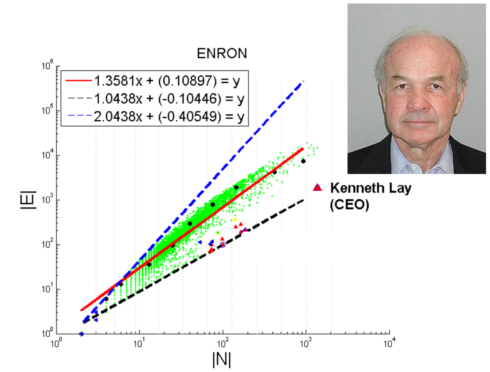
Tina Eliassi-Rad (Rutgers)

Christos Faloutsos (CMU)

SDM'12 Tutorial

Overview





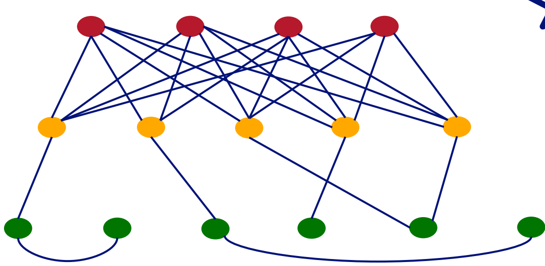
Features

Roles

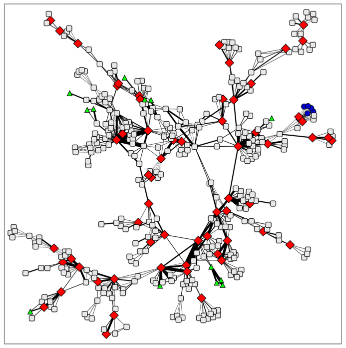
Anomalies

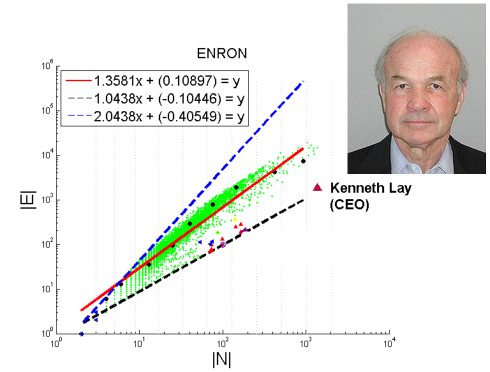
= rare roles

Patterns



Overview





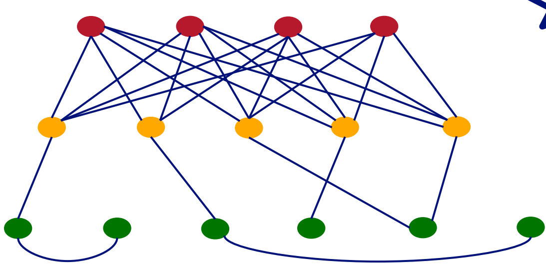
Features

Roles

Anomalies

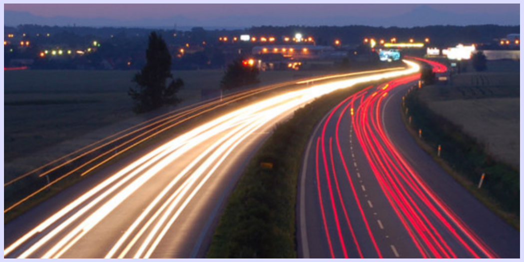
= rare roles

Patterns



Roadmap

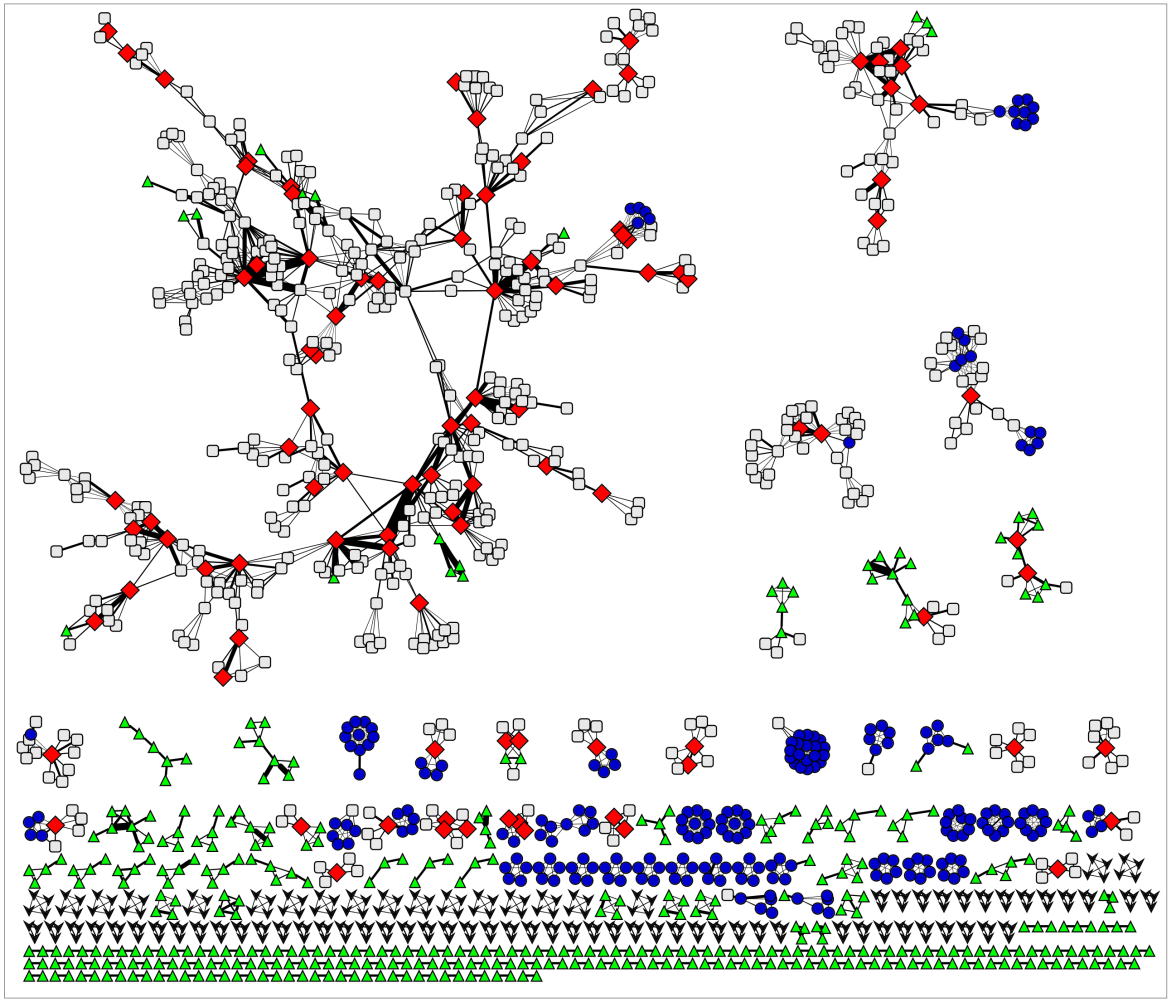
- What are roles
- Roles and communities
- Roles and equivalences (from sociology)
- Roles (from data mining)
- Summary



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

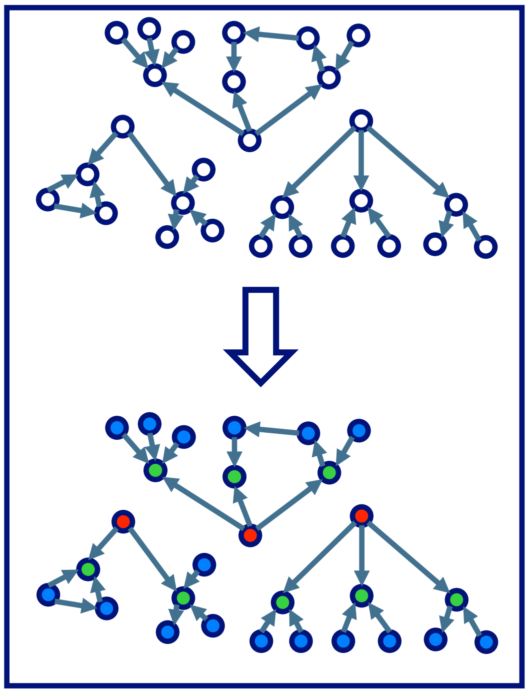
Example of Roles



- centers of stars
- members of cliques
- peripheral nodes

Why are roles important?

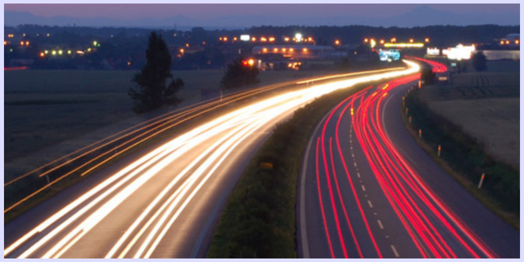
Role Discovery



- ✓ Automated discovery
- ✓ Behavioral roles
- ✓ Roles generalize

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
Identity resolution	Identify known individuals in a new network
Role transfer	Use knowledge of one network to make predictions in another
Network comparison	Determine network compatibility for knowledge transfer

Roadmap

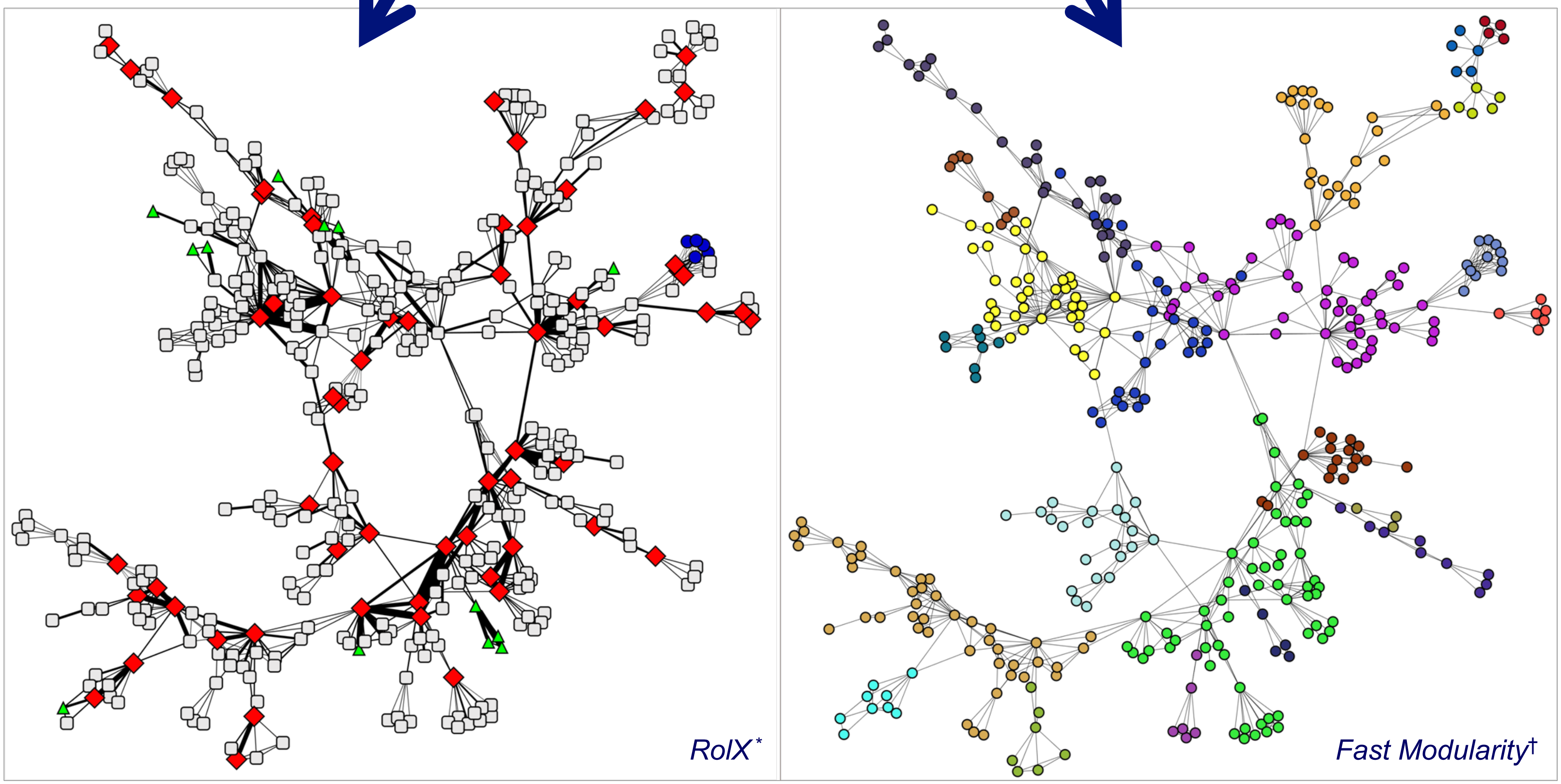


- What are roles
- Roles and communities
- Roles and equivalences (from sociology)
- Roles (from data mining)
- Summary

Roles and Communities

- Roles group nodes with similar structural properties
- Communities group nodes that are well-connected to each other
- Roles and communities are complementary

Roles and Communities



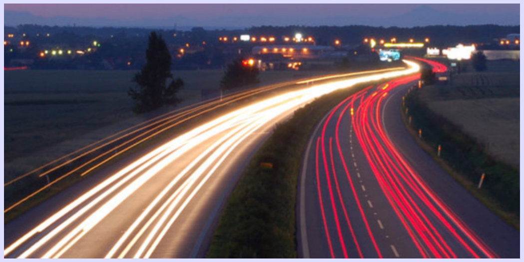
* Henderson, et al. 2012; † Clauset, et al. 2004

Roles and Communities

Consider the social network of a CS dept

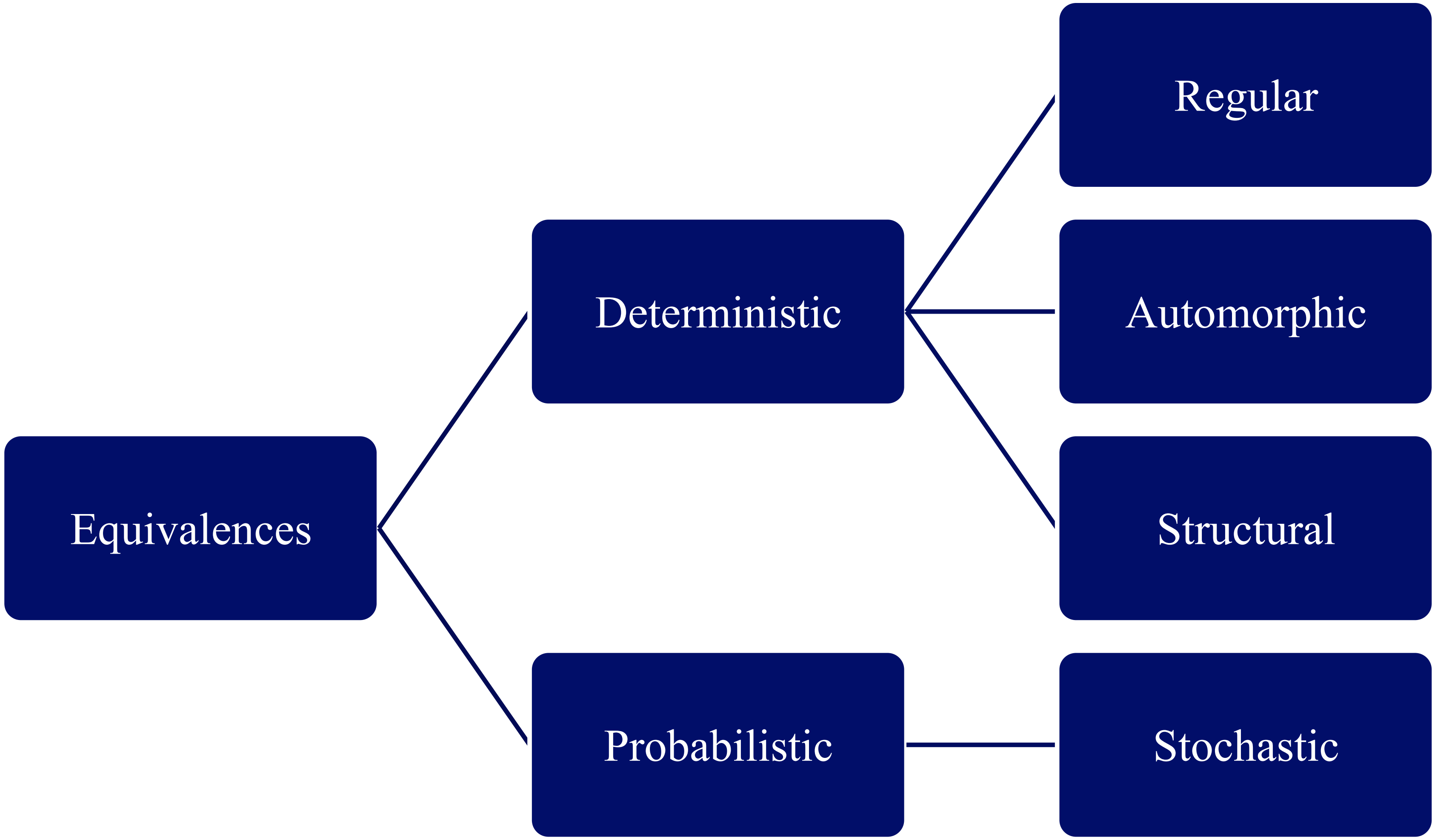
- Roles
  - Faculty
  - Staff
  - Students
  - …
- Communities
  - AI lab
  - Database lab
  - Architecture lab
  - …

Roadmap

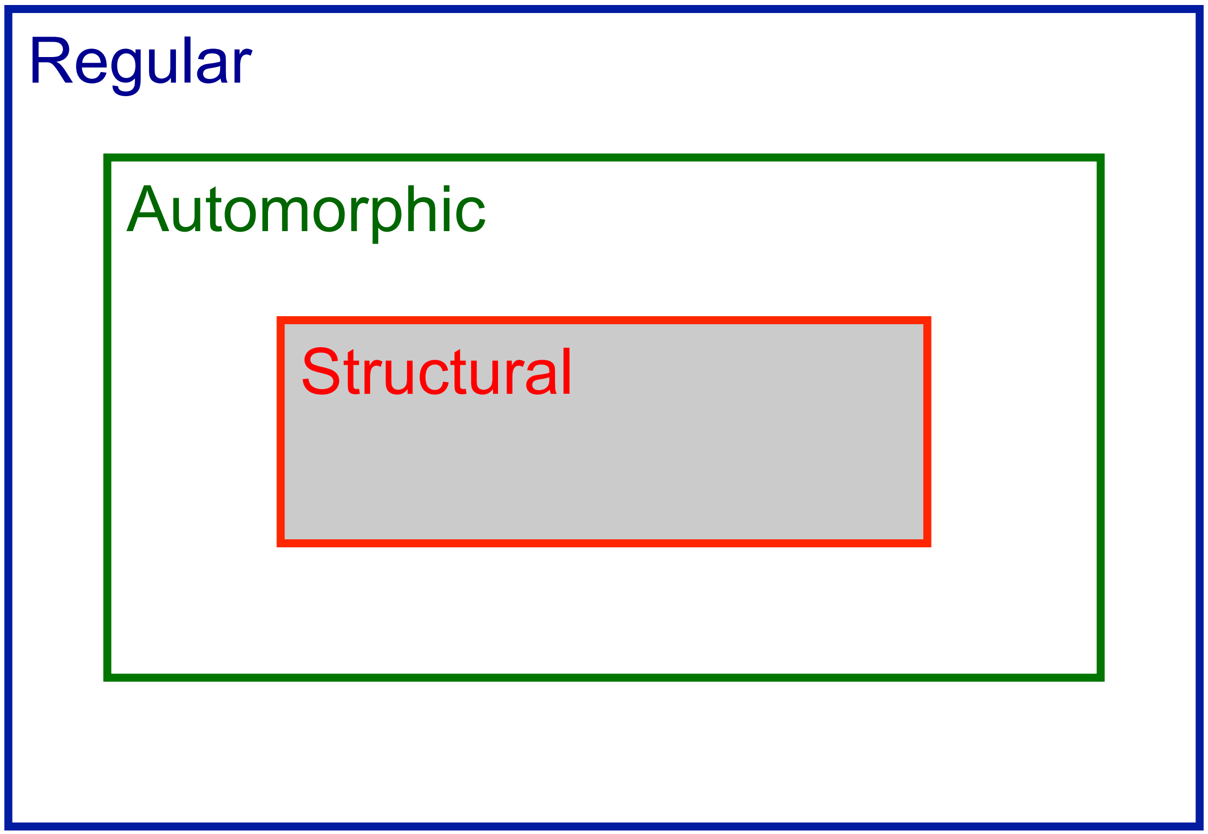


- What are roles
- Roles and communities
- Roles and equivalences (from sociology)
- Roles (from data mining)
- Summary

Equivalences

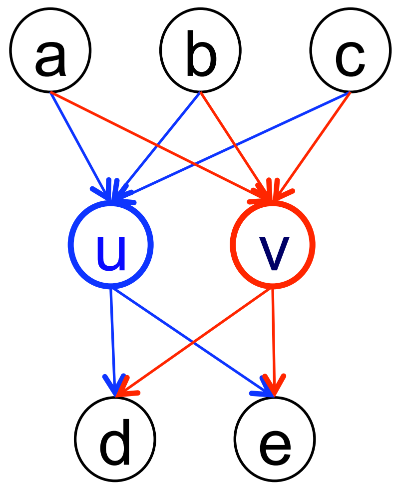


Deterministic Equivalences



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



Structural Equivalence: Algorithms

- CONCOR (CONvergence of iterated CORrelations) [Breiger et al. 1975]
- A hierarchical divisive approach
  1. Starting with the adjacency matrix, repeatedly calculate Pearson correlations between rows until the resultant correlation matrix consists of +1 and -1 entries
  2. Split the last correlation matrix into two structurally equivalent submatrices (a.k.a. blocks): one with +1 entries, another with -1 entries
- Successive split can be applied to submatrices in order to produce a hierarchy (where every node has a unique position)

Structural Equivalence: Algorithms

- STRUCUTRE [Burt 1976]
- A hierarchical agglomerative approach
  1. For each node *i*, create its ID vector by concatenating its row and column vectors from the adjacency matrix
  2. For every pair of nodes $\langle i, j \rangle$, measure the square root of sum of squared differences between the corresponding entries in their ID vectors
  3. Merge entries in hierarchical fashion as long as their difference is less than some threshold α

Structural Equivalences: Algorithms

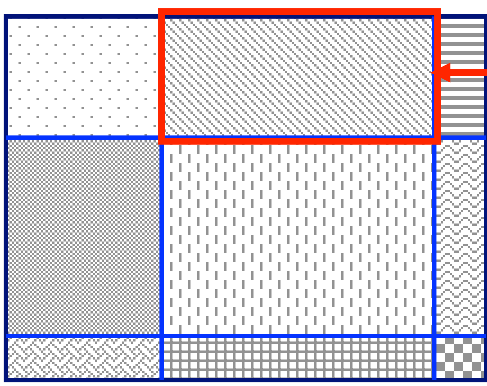
- Combinatorial optimization approaches
  - Numerical optimization with tabu search [UCINET]
  - Local optimization [Pajek]
- Partition the sociomatrices into blocks based on a cost function that minimizes the sum of within block variances
  - Basically, minimize the sum of code cost within each block

Cross-Associations (XA)

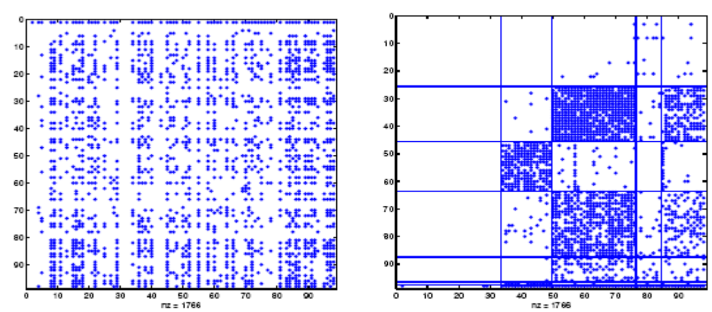
- [Chakrabarti+, KDD 2004]
- Minimize total encoding cost of the adjacency matrix

$$\overbrace{\sum_i \left((n_i^1+n_i^0)\times H(p_i^1)\right)}^{\text{Code Cost}} + \overbrace{\sum_i \left(\text{cost of describing } n_i^1, n_i^0 \text{ and groups}\right)}^{\text{Description Cost}}$$

Binary Matrix

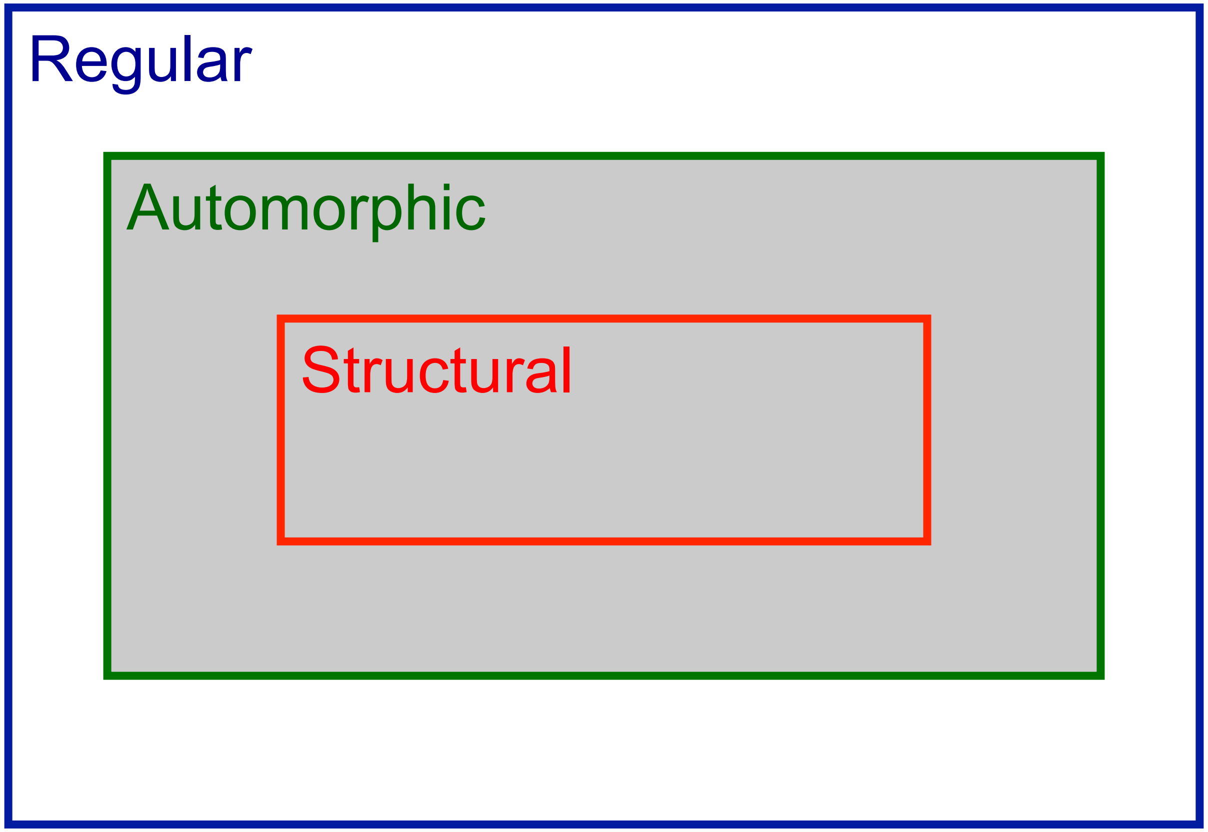


$p_i^1 = n_i^1 / (n_i^1 + n_i^0)$



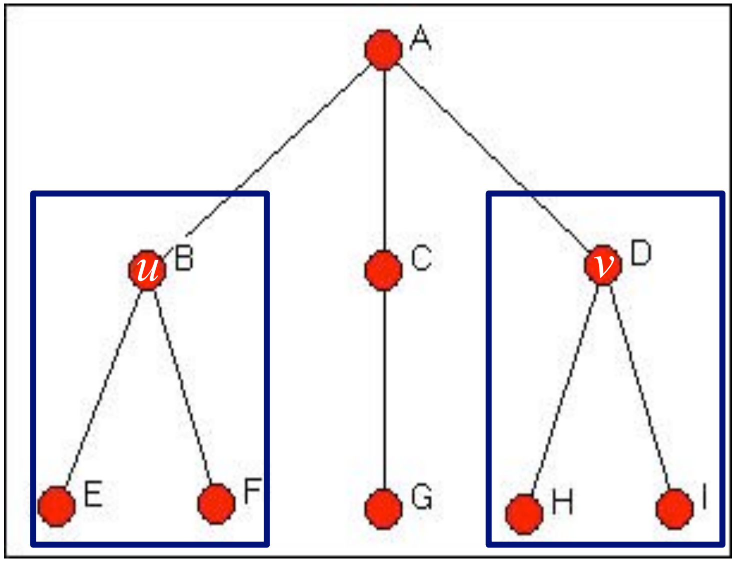
(a) before (b) after

Deterministic Equivalences



Automorphic Equivalence

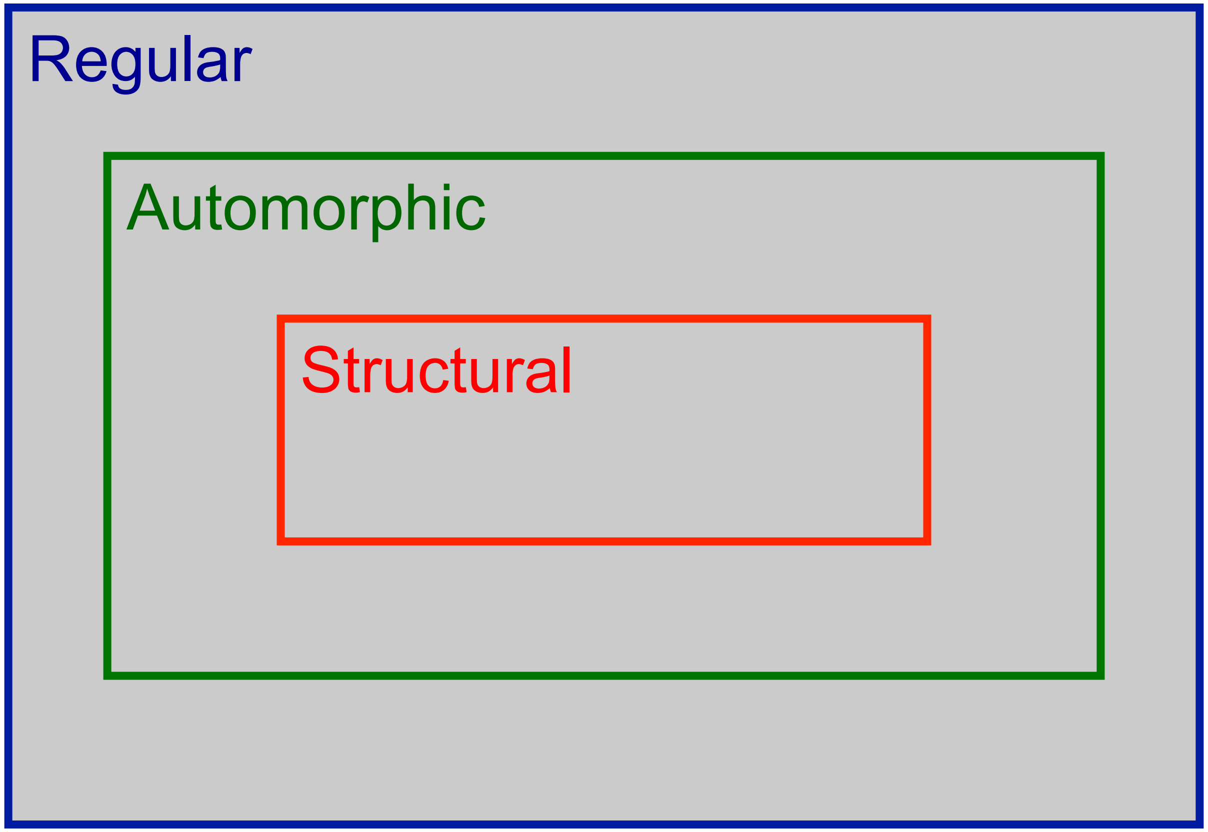
- [Borgatti, et al. 1992; Sparrow 1993]
- Two nodes *u* and *v* are automorphically equivalent if all the nodes can be relabeled to form an isomorphic graph with the labels of *u* and *v* interchanged
  - Swapping *u* and *v* (possibly along with their neighbors) does not change graph distances
- Two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



Automorphic Equivalence: Algorithms

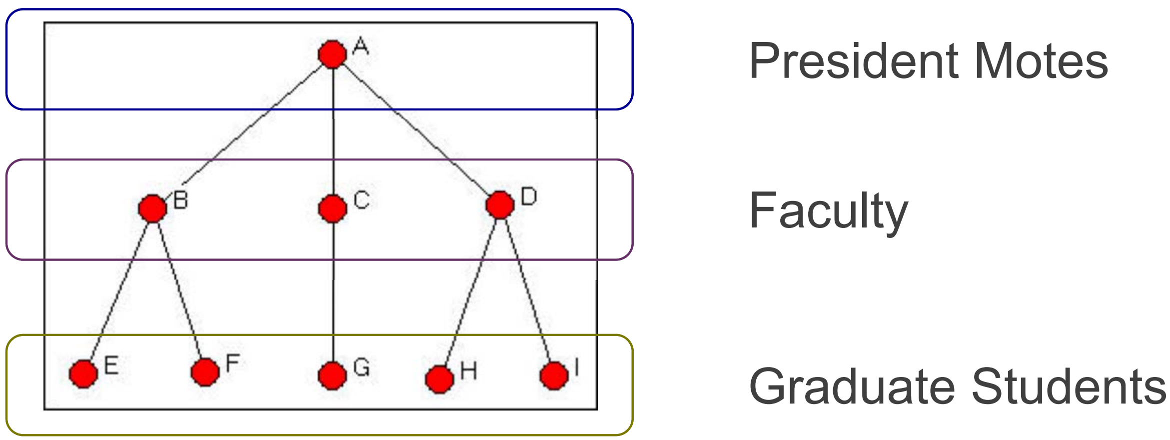
- Sparrow (1993) proposed an algorithm that scales linearly to the number of edges
- Use numerical signatures on degree sequences of neighborhoods
- Numerical signatures use a unique transcendental number like π, which is independent of any permutation of nodes
- Suppose node i has the following degree sequence: 1, 1, 5, 6, and 9. Then its signature is $S_{i,1} = (1+\pi)(1+\pi)\,(5+\pi)\,(6+\pi)\,(9+\pi)$
- The signature for node i at k+1 hops is $S_{i,(k+1)} = \Pi(S_{i,k} + \pi)$
- To find automorphic equivalence, simply compare numerical signatures of nodes

Deterministic Equivalences



Regular Equivalence

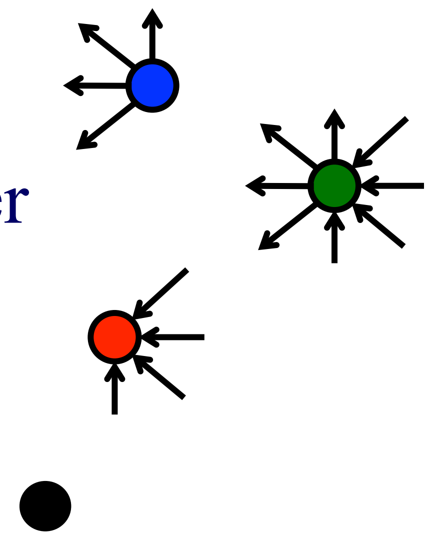
- [Everett & Borgatti, 1992]
- Two nodes *u* and *v* are regularly equivalent *if* they are equally related to equivalent others



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/)

Regular Equivalence
(continued)

- Basic roles of nodes
  - source
  - repeater
  - sink
  - isolate



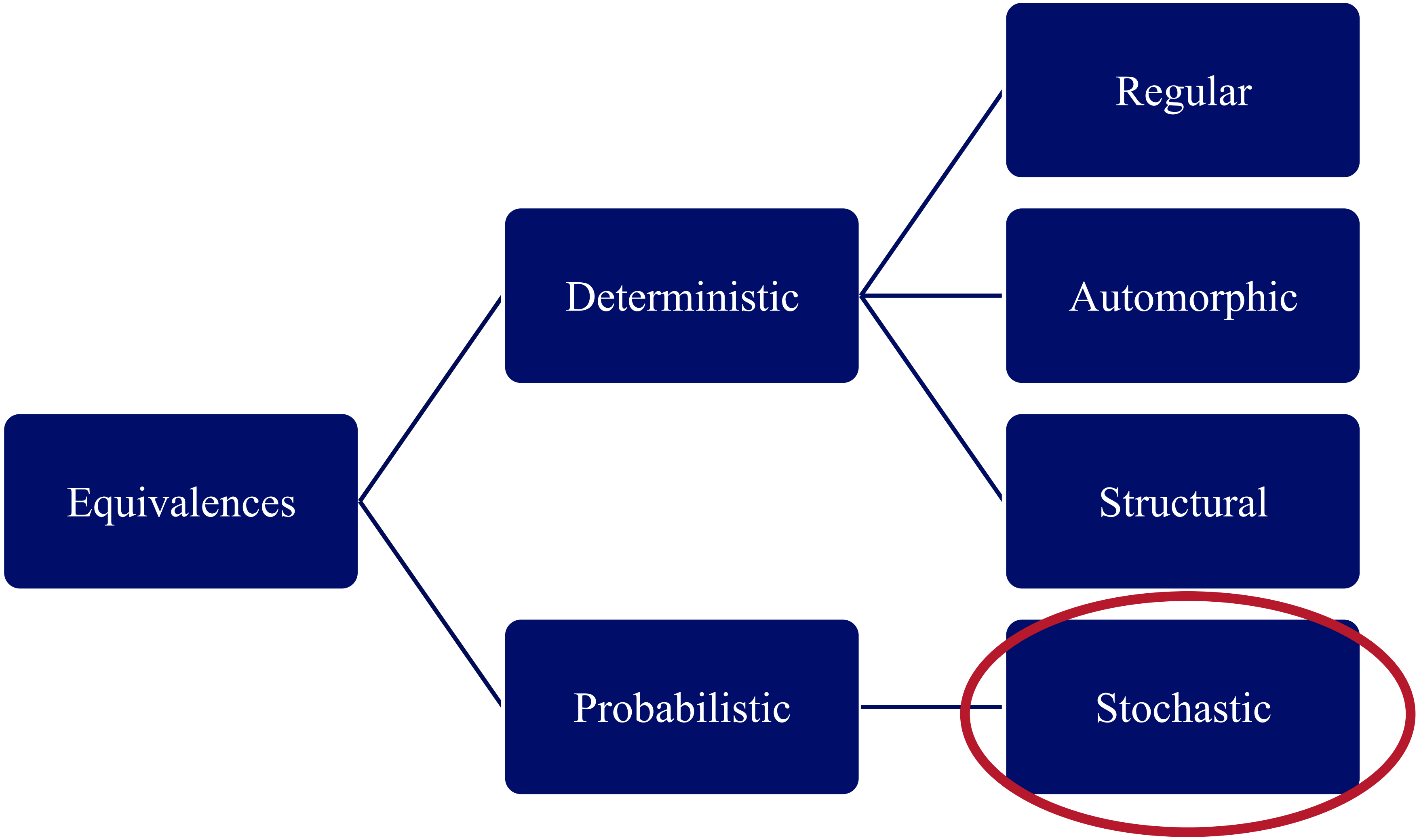
Regular Equivalence
(continued)

- Based solely on the social roles of neighbors
- Interested in
  - Which nodes fall in which social roles?
  - How do social roles relate to each other?
- Hard partitioning of the graph into social roles
- A given graph can have more than one valid regular equivalence set
- Exact regular equivalences can be rare in large graphs

Regular Equivalence: Algorithms

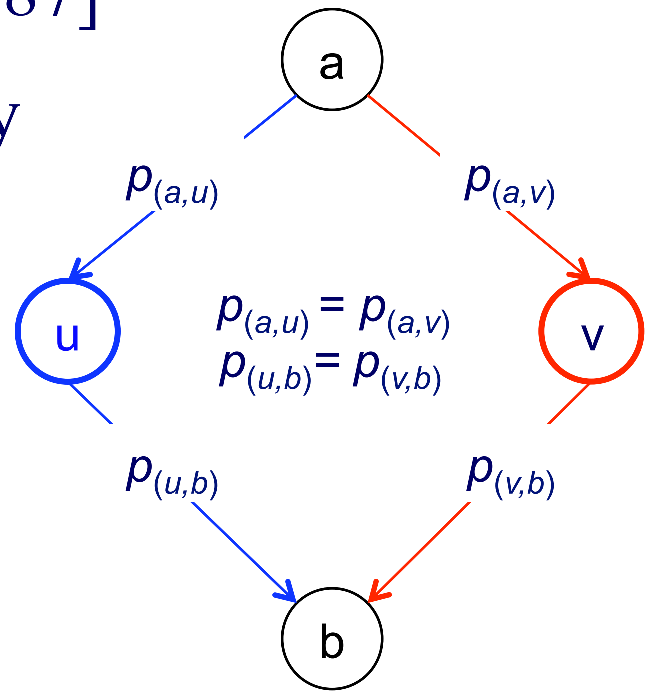
- Many algorithms exist here
- Basic notion
  - Profile each node's neighborhood by the presence of nodes of other "types"
  - Nodes are regularly equivalent to the extent that they have similar "types" of other nodes at similar distances in their neighborhoods

Equivalences



Stochastic Equivalence

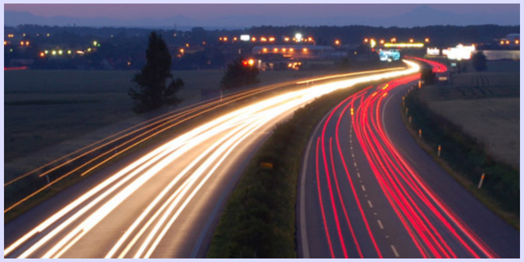
- [Holland, et al. 1983; Wasserman & Anderson, 1987]
- Two nodes are stochastically equivalent if they are "exchangeable" w.r.t. a probability distribution
- Similar to structural equivalence but probabilistic



Stochastic Equivalence: Algorithms

- 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] equivalences (a.k.a. "positions")
  - Parametric vs. non-parametric models

Roadmap

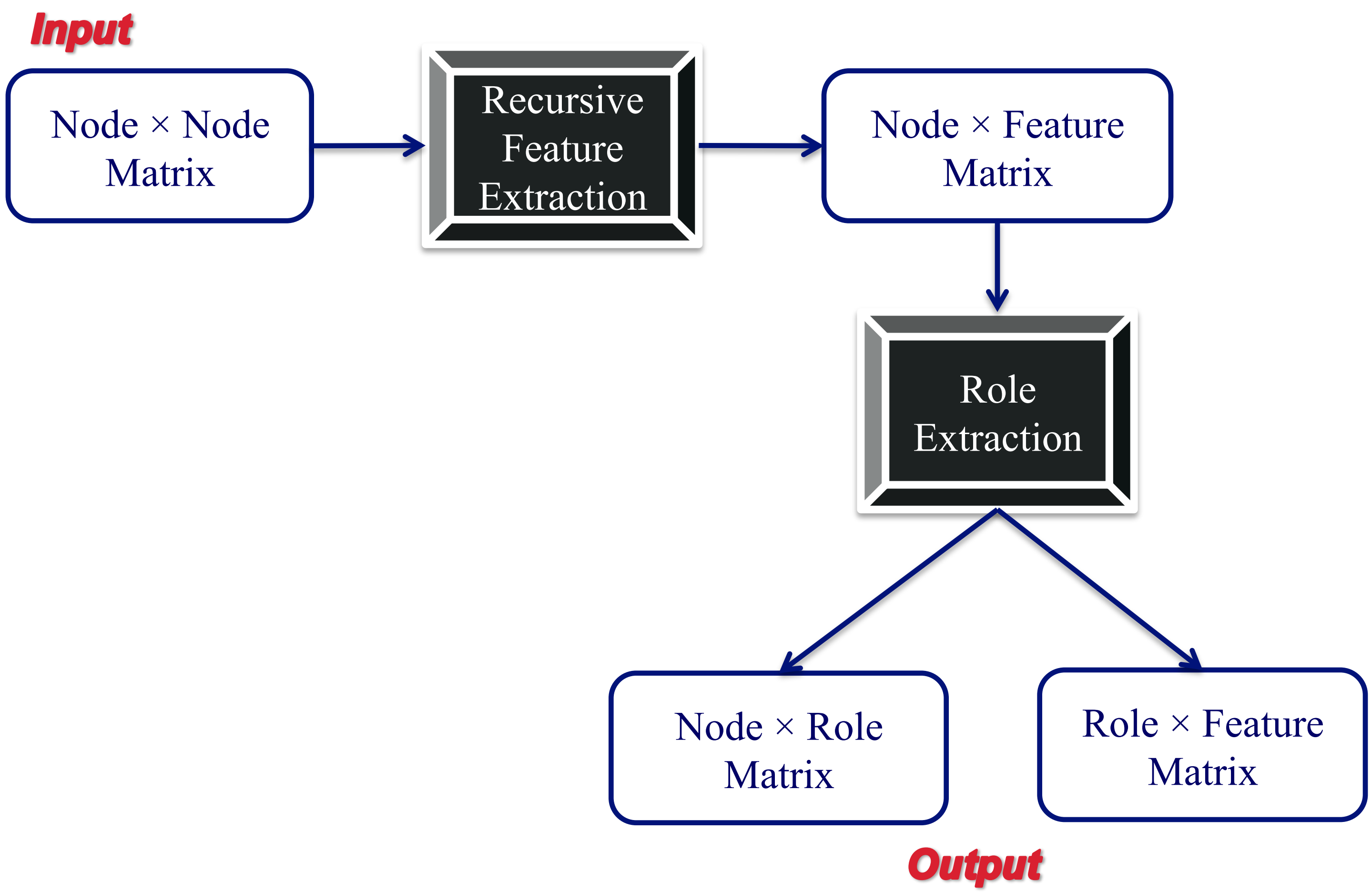


- What are roles
- Roles and communities
- Roles and equivalences (from sociology)
- Roles (from data mining)
- Summary

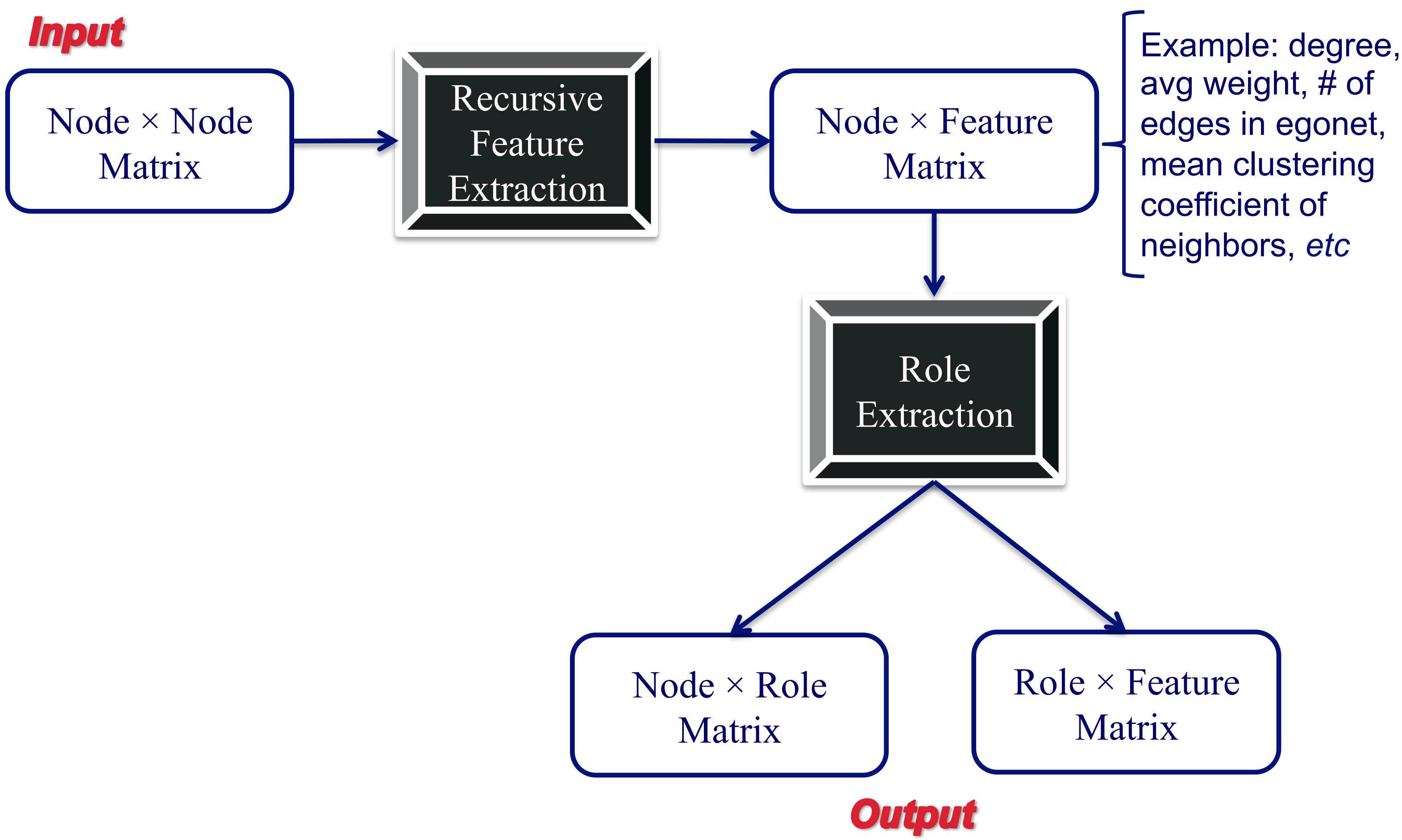
RolX: Role eXtraction

- Introduced by Henderson, et al. 2011b
- Automatically extracts the underlying roles in a network
  - No prior knowledge required
- Assigns a mixed-membership of roles to each node
- Scales linearly on the number of edges

RolX: Flowchart

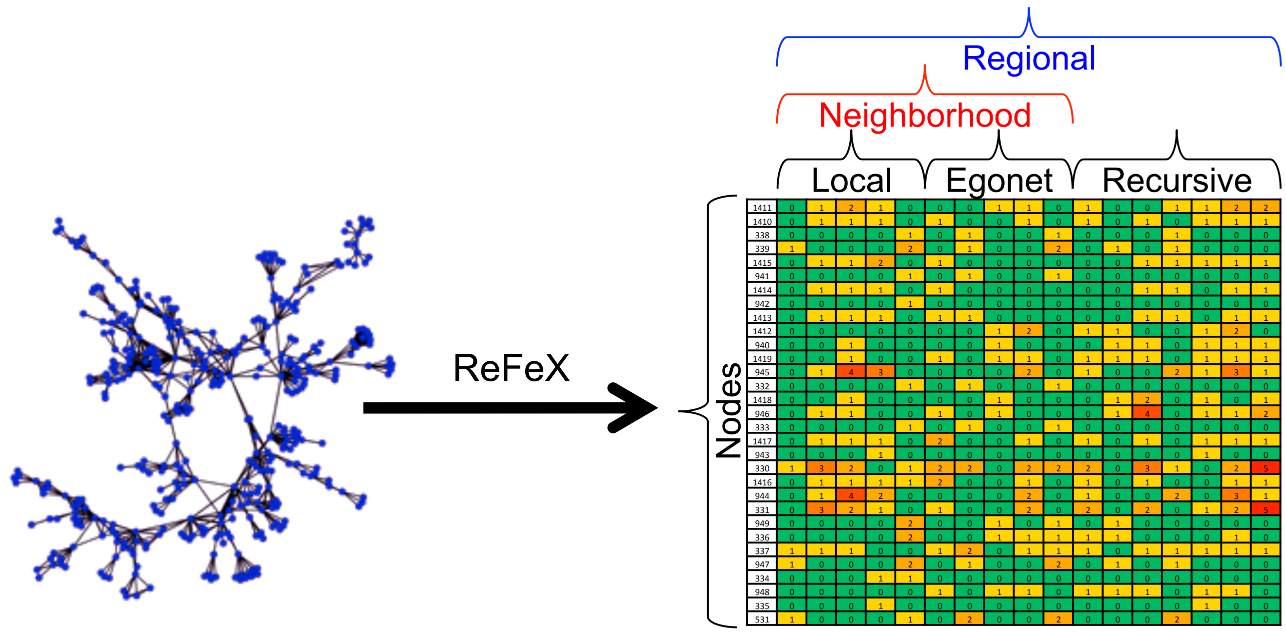


RolX: Flowchart



Recursive Feature Extraction

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



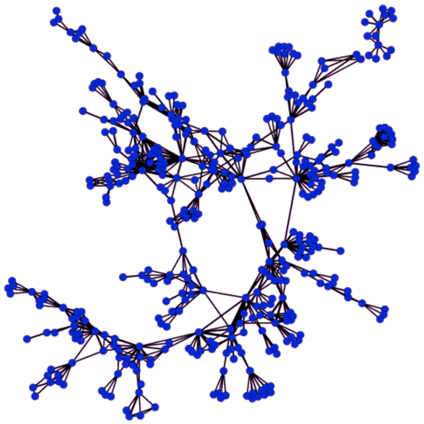
- Neighborhood features: What is your connectivity pattern?
- Recursive Features: To what *kinds* of nodes are you connected?

Propositionalisation (PROP)

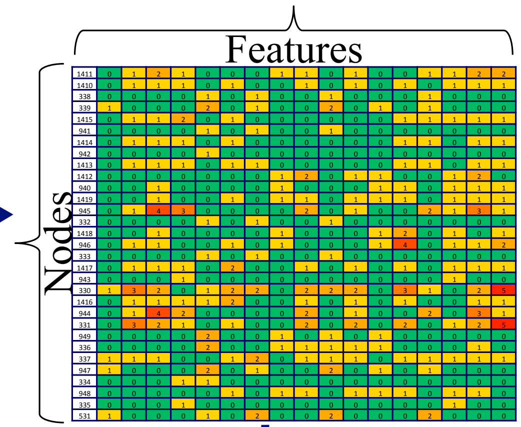
- [Knobbe, et al. 2001; Neville, et al. 2003; Krogel, et al. 2003]
- From multi-relational data mining with roots in Inductive Logic Programming (ILP)
- Summarizes a multi-relational dataset (stored in multiple tables) into a propositional dataset (stored in a single “target” table)
- Derived attribute-value features describe properties of individuals
- Related more to recursive structural features than structural roles

Role Extraction

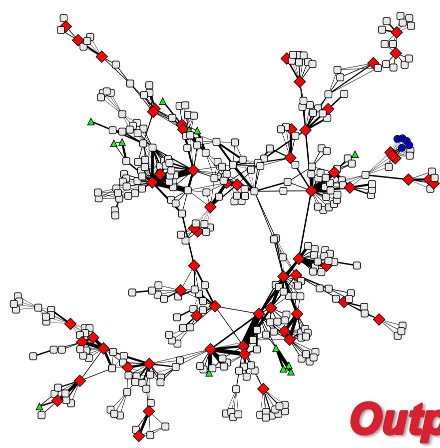
Input



Recursively extract features

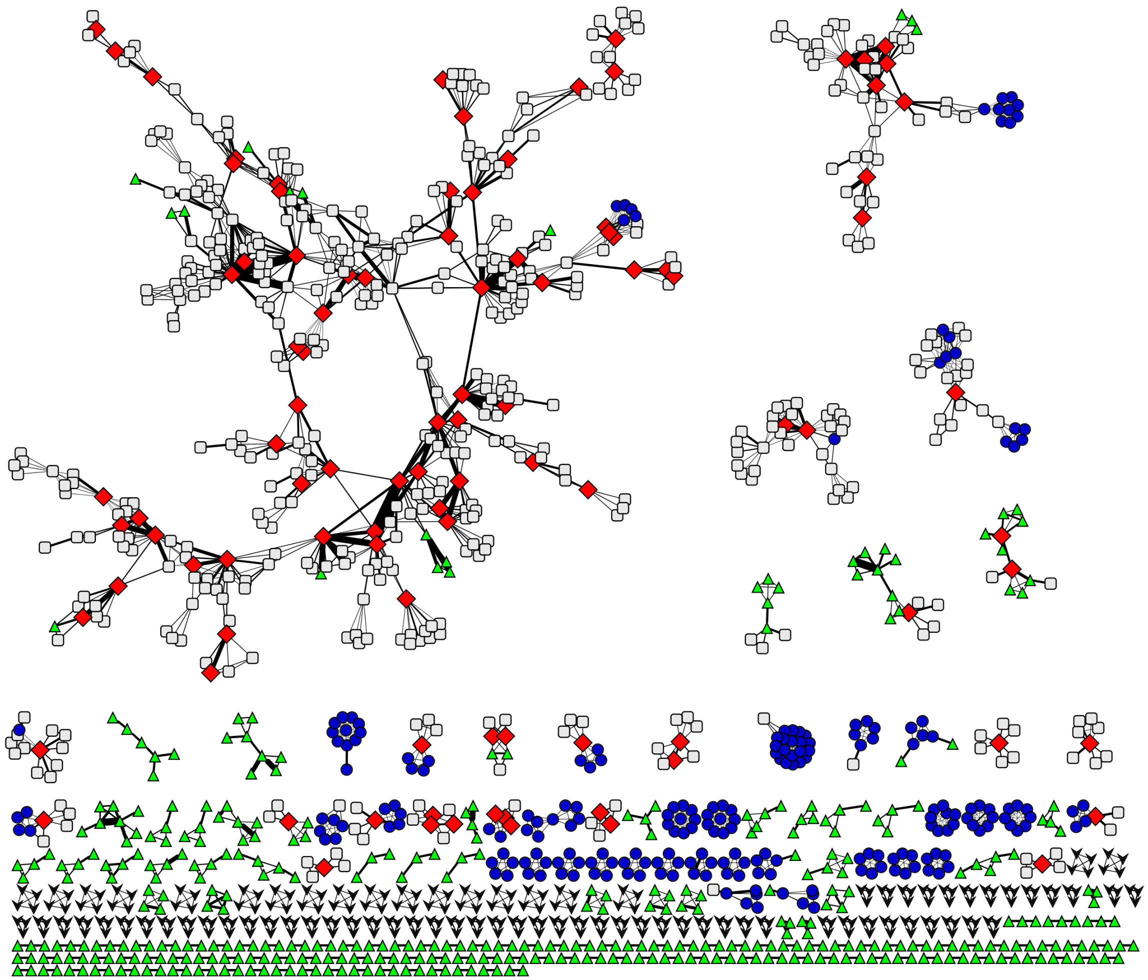


Automatically factorize roles



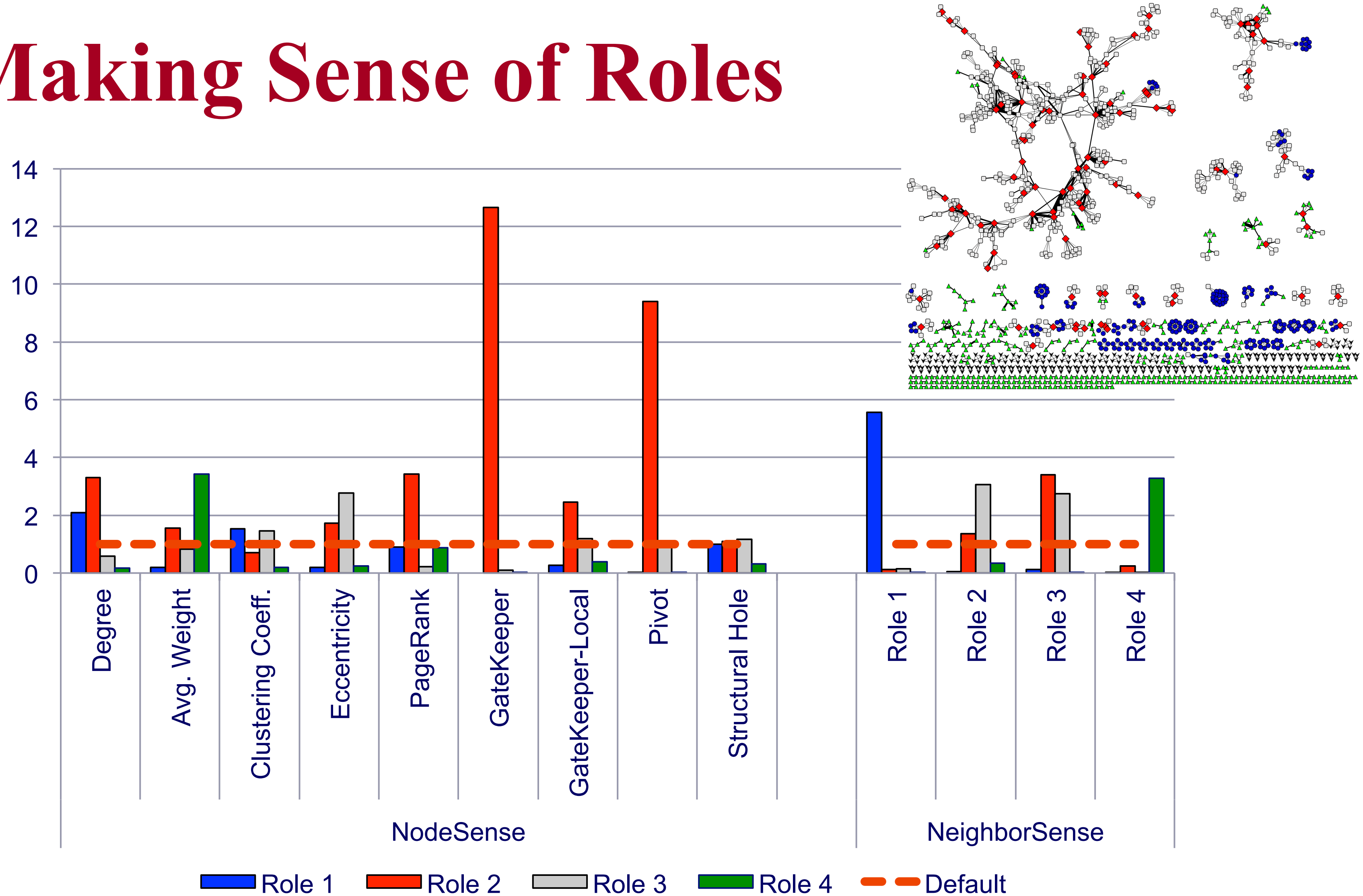
Output

Automatically Discovered Roles



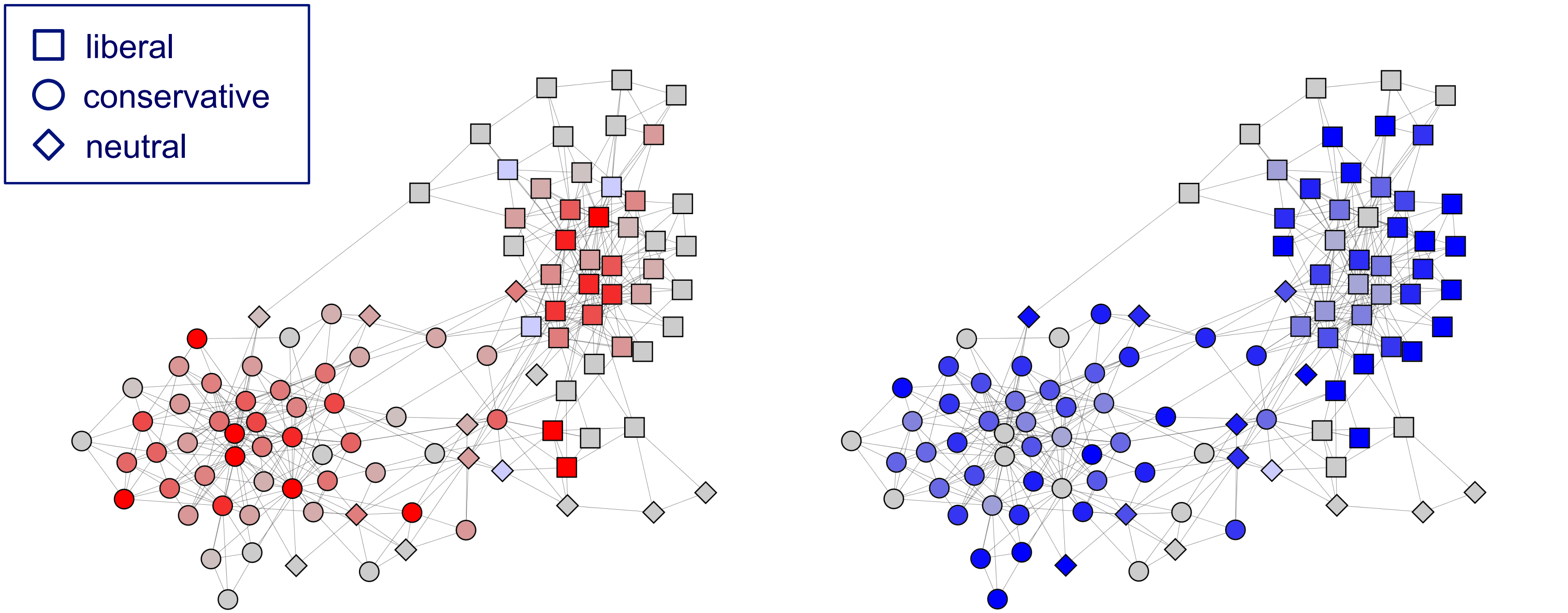
Network Science Co-authorship Graph [Newman 2006]

Making Sense of Roles



cliquey *bridge* *periphery* *isolated*

Mixed-Membership over Roles

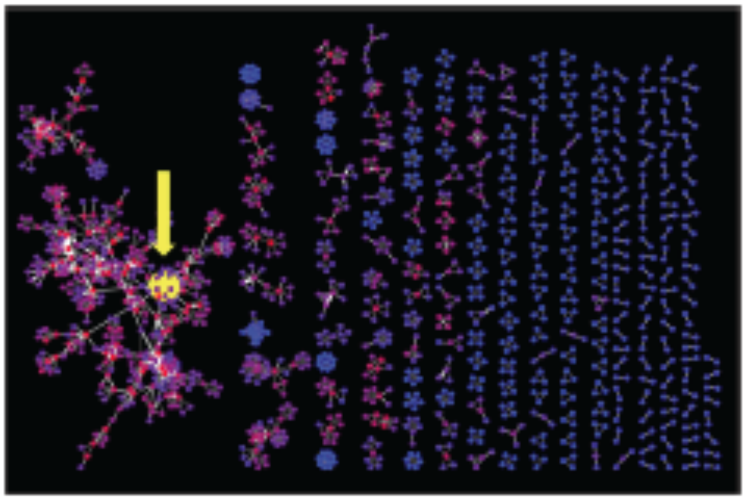


Bright red nodes are locally central nodes

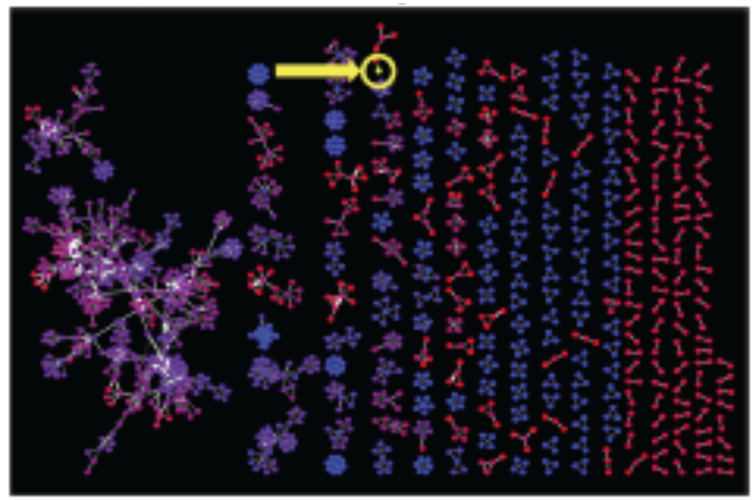
Bright blue nodes are peripheral nodes

Amazon Political Books Co-purchasing Network
[V. Krebs 2000]

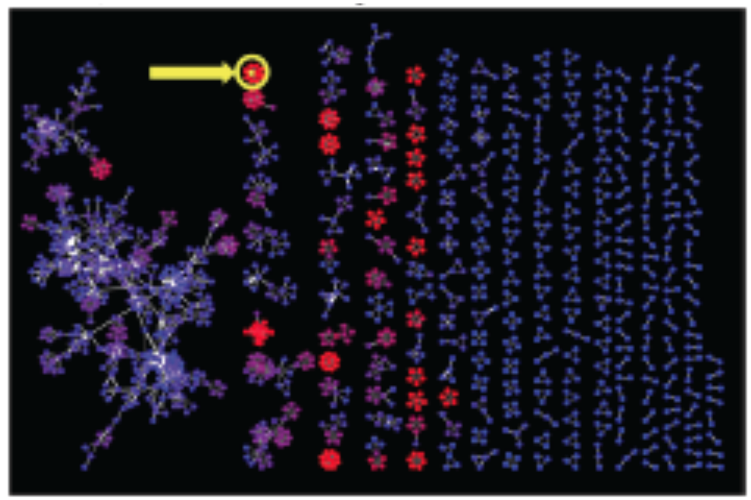
Role Query



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

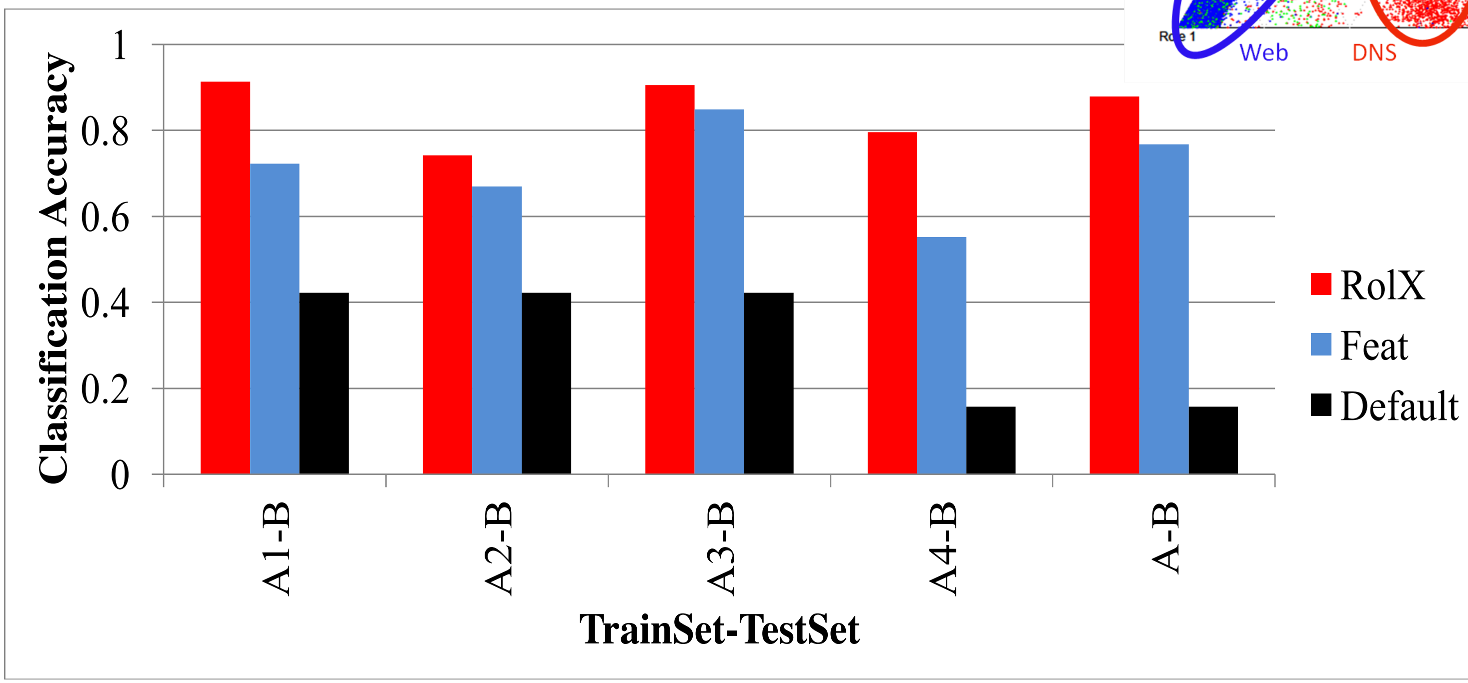


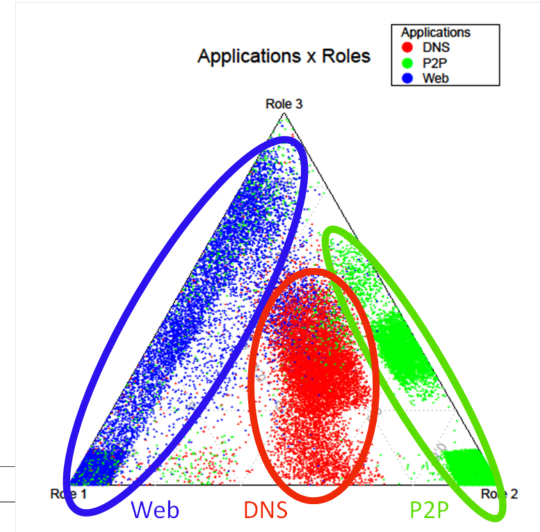
Node Similarity for J. Rinzel (*isolate*)



Node Similarity for F. Robert (*cliquey*)

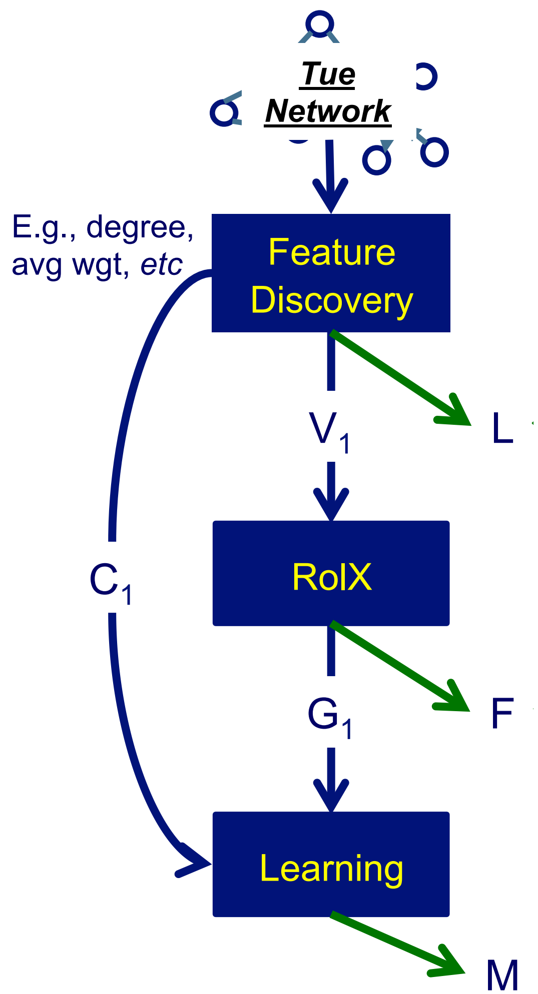
Roles Generalize across Disjoint Networks





Roles Generalize across Networks

Discovery Stage



V: (node × feature) matrix

G: (node × role) matrix

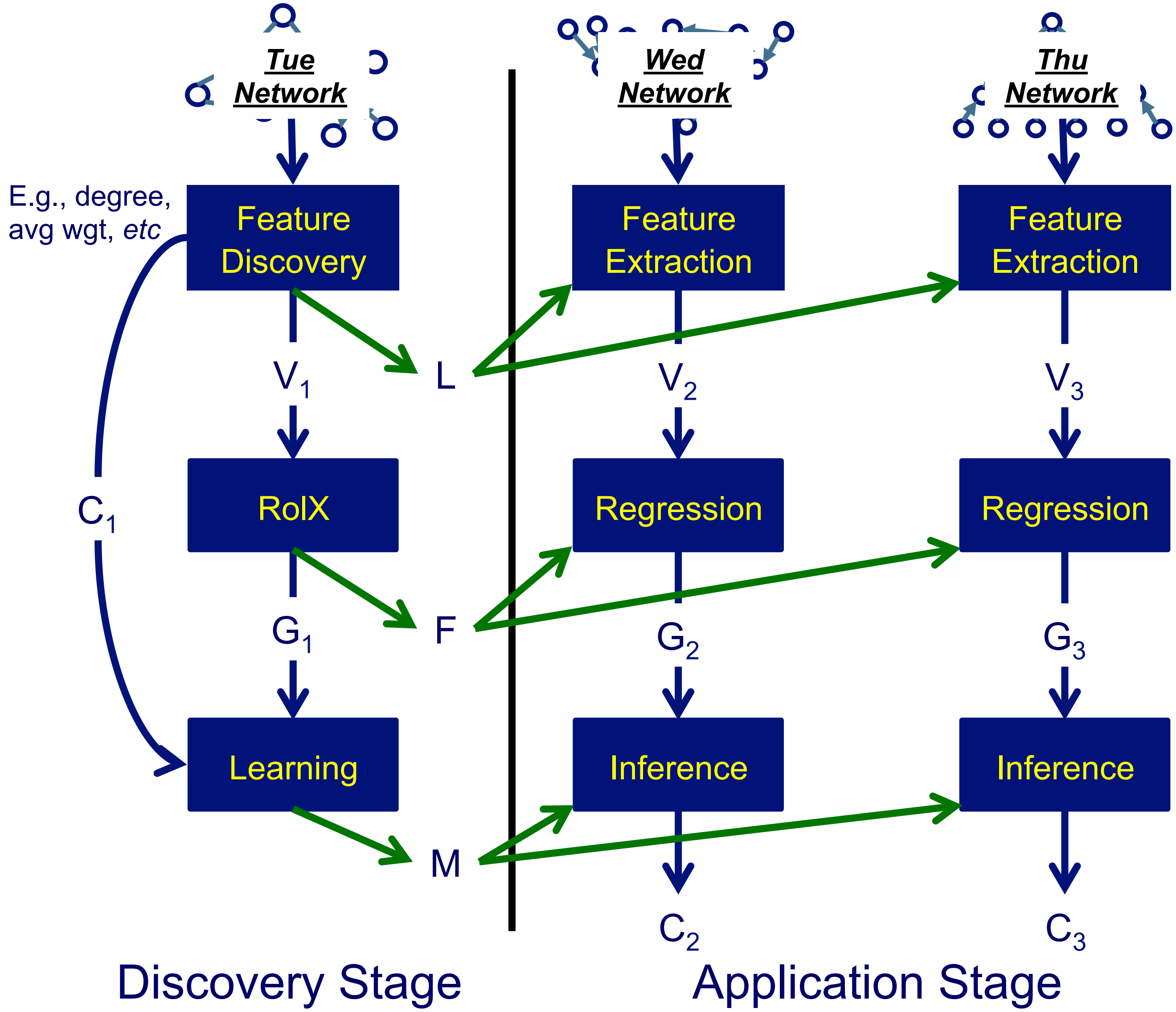
F: (role × feature) matrix

L: List of feature names

C: Class labels

M: model

Roles Generalize across Networks



V: (node × feature) matrix

G: (node × role) matrix

F: (role × feature) matrix

L: List of feature names

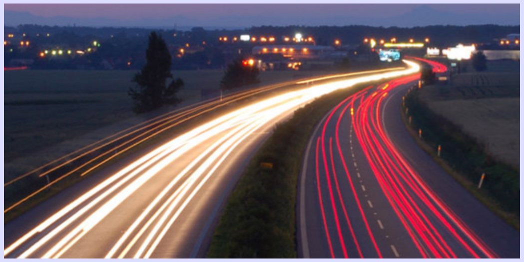
C: Class labels

M: model

Roles: Regular Equivalence vs. RolX

	RolX	Regular Equivalence
Mixed-membership over roles	✓	
Fully automatic	✓	
Uses structural features	✓	
Uses structure	✓	✓
Generalizable across disjoint networks	✓	?
Scalable (linear on # of edges)	✓	

Roadmap



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

Summary

- Roles
  - Structural behavior (“function”) of nodes
  - 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, node dynamics, *etc*

Acknowledgement

- LLNL: Brian Gallagher, Keith Henderson
- IBM Watson: Hanghang Tong
- Google: Sugato Basu
- CMU: Leman Akoglu, Danai Koutra, Lei Li

Thanks to: LLNL, NSF, IARPA.

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.
- K. Henderson, B. Gallagher, T. Eliassi-Rad, H. Tong, S. Basu, L. Akoglu, D. Koutra, L. Li, C. Faloutsos: RolX: Structural Role Extraction & Mining in Large Graphs. Technical Report, Lawrence Livermore National Laboratory, Livermore, CA, 2011.
- Ruoming Jin, Victor E. Lee, Hui Hong: Axiomatic ranking of network role similarity. KDD 2011: 922-930.

References

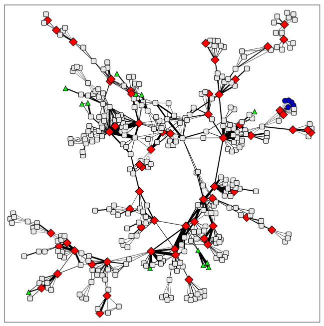
Community Discovery

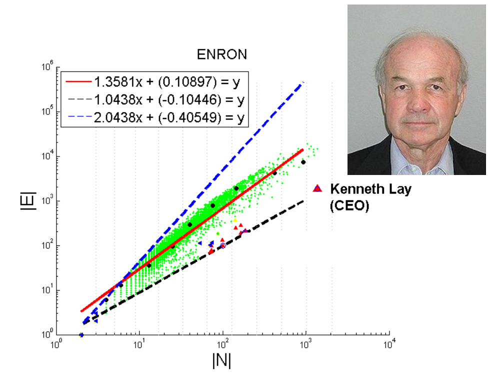
- A. Clauset, M.E.J. Newman, C. Moore: Finding Community Structure in Very Large Networks. Phys. Rev. E., 70:066111, 2004.
- M.E.J. Newman: Finding Community Structure in Networks Using the Eigenvectors of Matrices. Phys. Rev. E., 74:036104, 2006.

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.

Back to Overview





Features

Roles

Anomalies

= rare roles

Patterns