Northeastern University
Network Science Institute

THE REASONABLE EFFECTIVENESS OF ROLES IN NETWORKS

Tina Eliassi-Rad
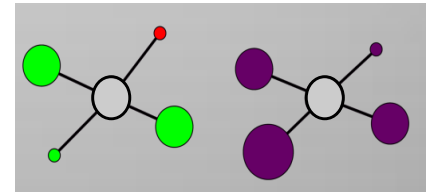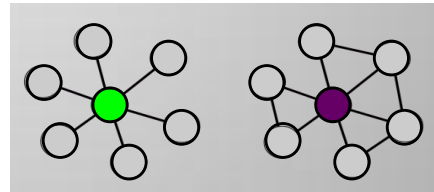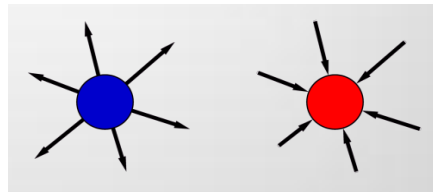
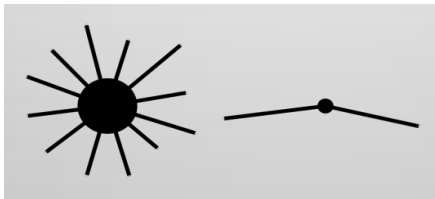tina@eliassi.org   @tinaeliassi

What are Roles?

- Functions of nodes in the network

  - Similar to functional roles of species in ecosystems

- Roles are defined in terms of structural behaviors

  - What is your connectivity pattern?

  - To what kinds of individuals are you connected?

Example of Roles in an IP × IP Network

Web	DNS	Peer-to-Peer
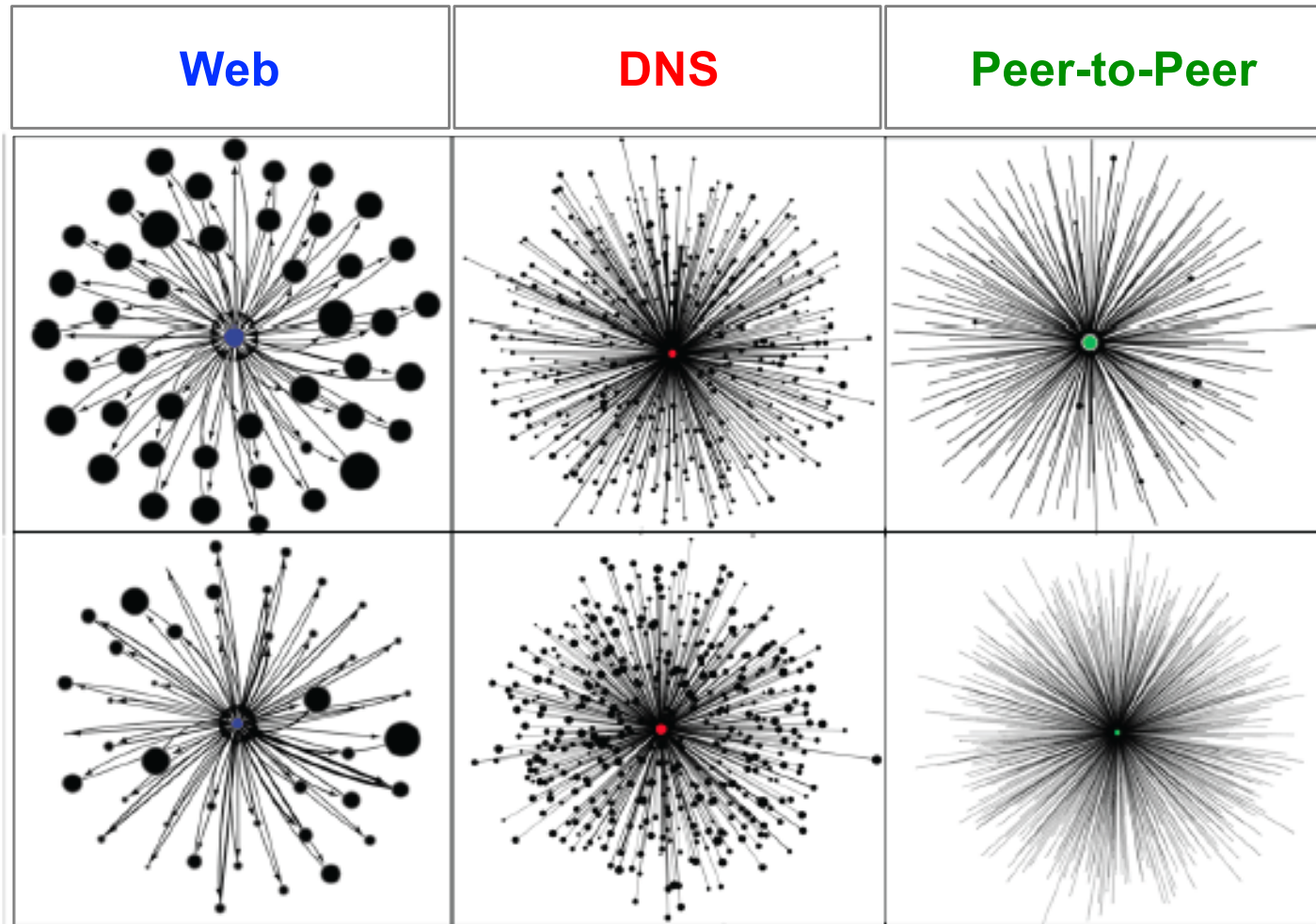
Node sizes indicate communication volume relative to the central node in each frame.

The types of neighbors that are connected to a given host are indicators of the host's role.

Each Node has a <u>Mixture</u> of Roles
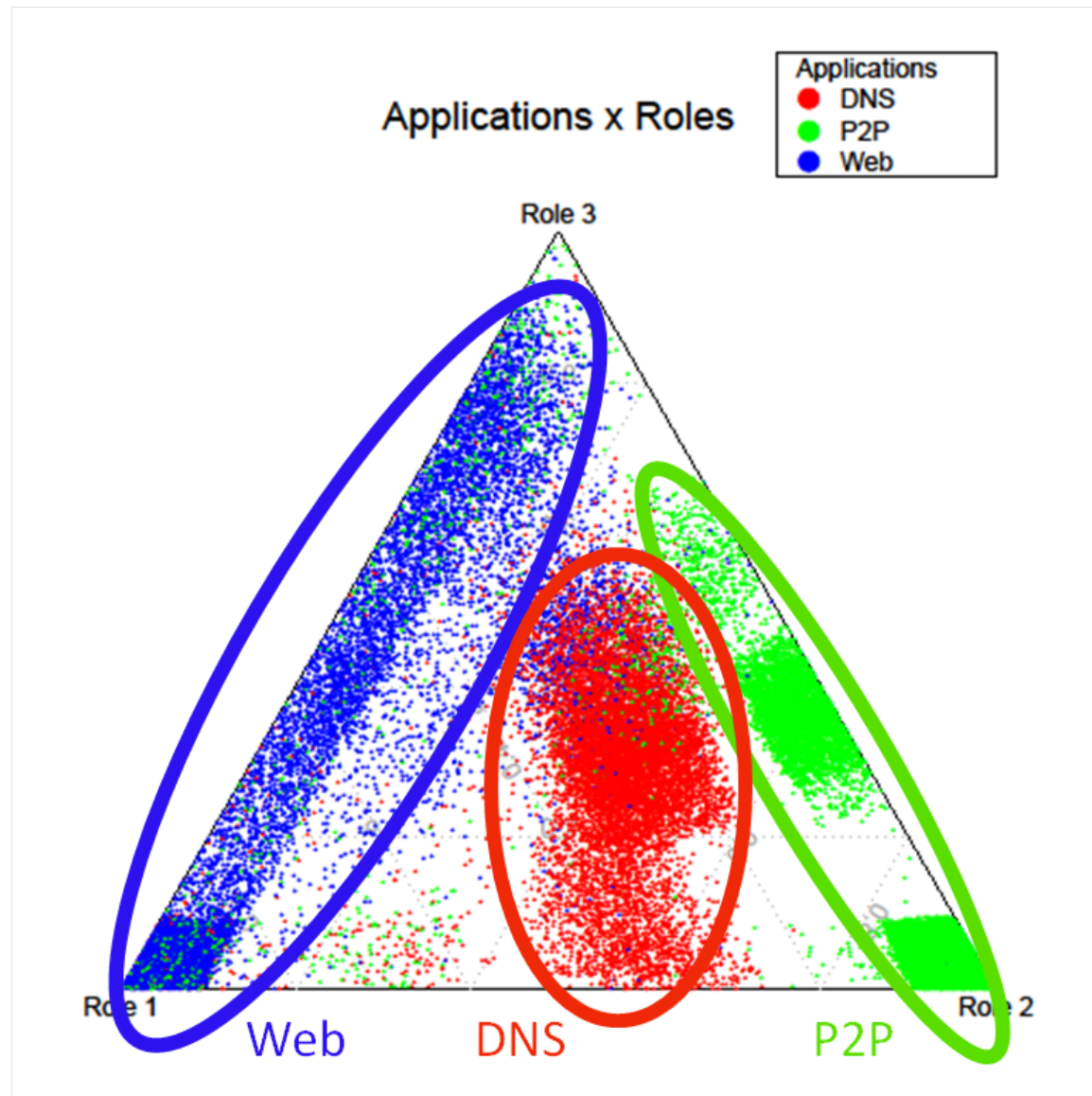
Research Questions

1. How are roles different from communities and from positions/equivalences (from sociology)?

2. Given a network, how can we automatically discover roles of nodes?

3. How can we make sense of these roles?

4. Are there good features that we can extract for nodes that indicate role-membership?

5. What are the applications in which these discovered roles can be effectively used?

Roles & Communities are Complementary

Roles
(similar structural properties)
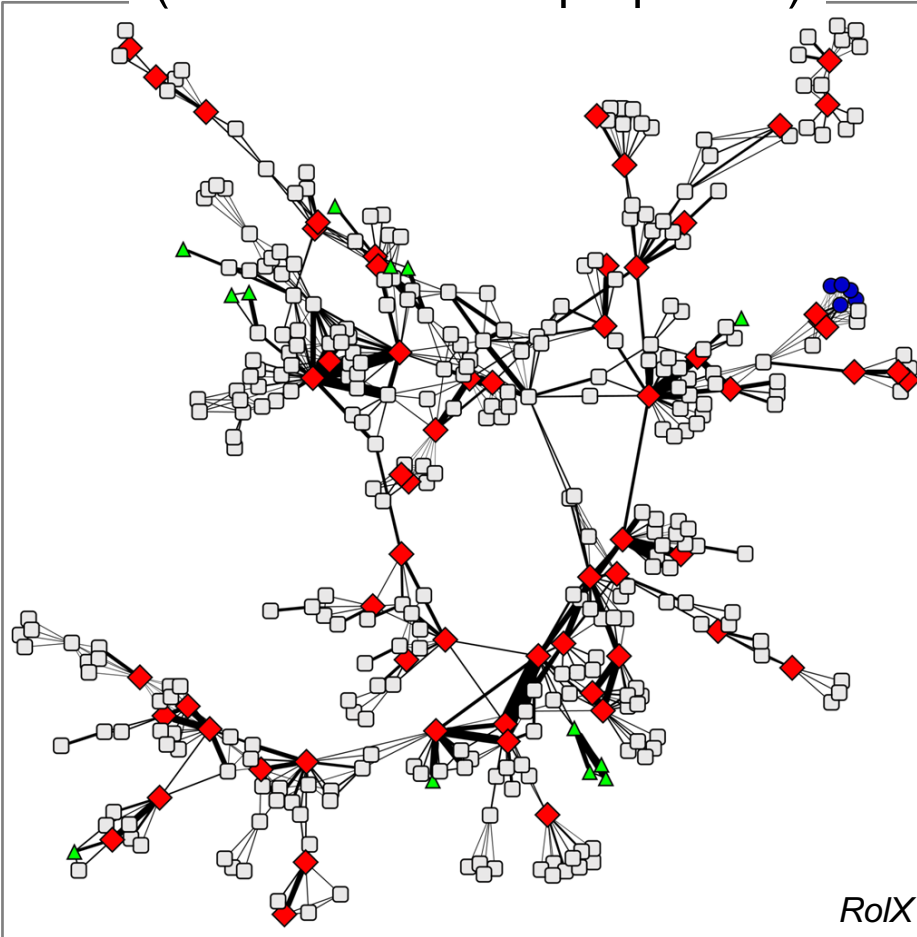
Communities
(well-connectedness)

RolX[*]

Fast Modularity[†]
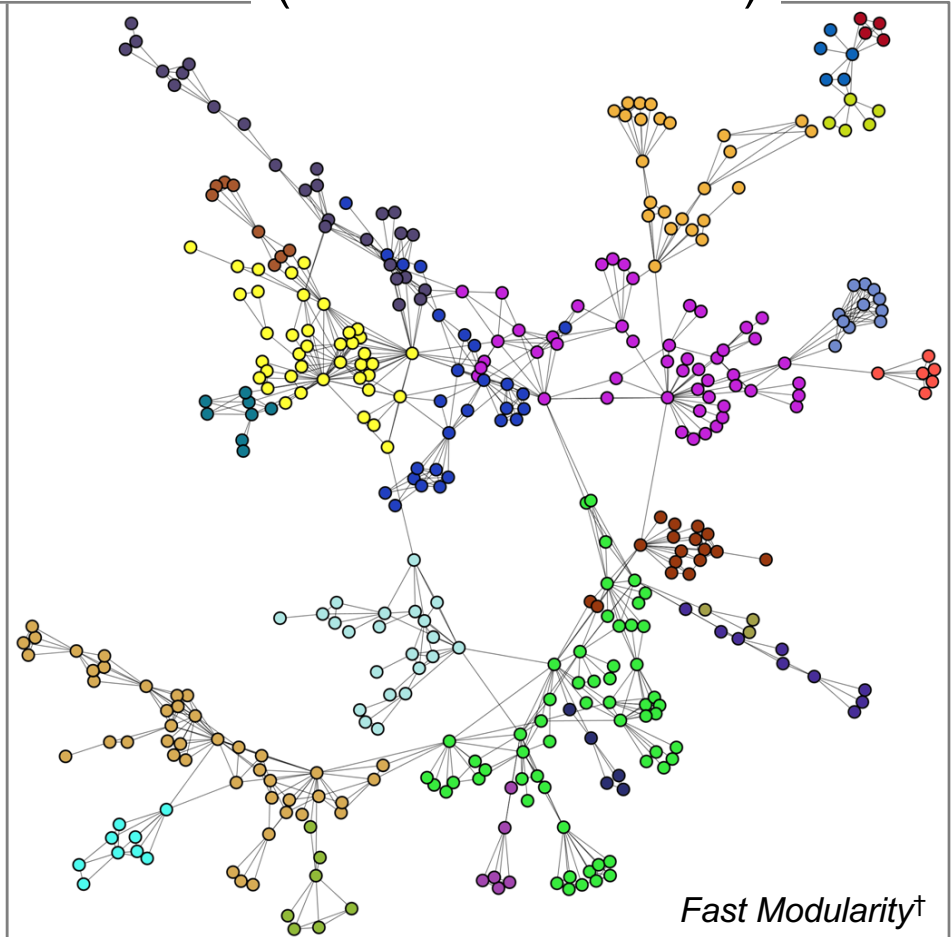
[*] Henderson, et al. 2012; [†] Clauset, et al. 2004

Roles are Similar to Positions from Sociology

- Two nodes with the same position are in an equivalence relation

- Equivalence, Q, is any relation that satisfies these three conditions:

    - Transitivity: $(a,b), (b,c) \in Q \Rightarrow (a,c) \in Q$

    - Symmetry: $(a, b) \in Q$ if and only if $(b, a) \in Q$

    - Reflexivity: $(a, a) \in Q$

Taxonomy of Equivalences from Sociology
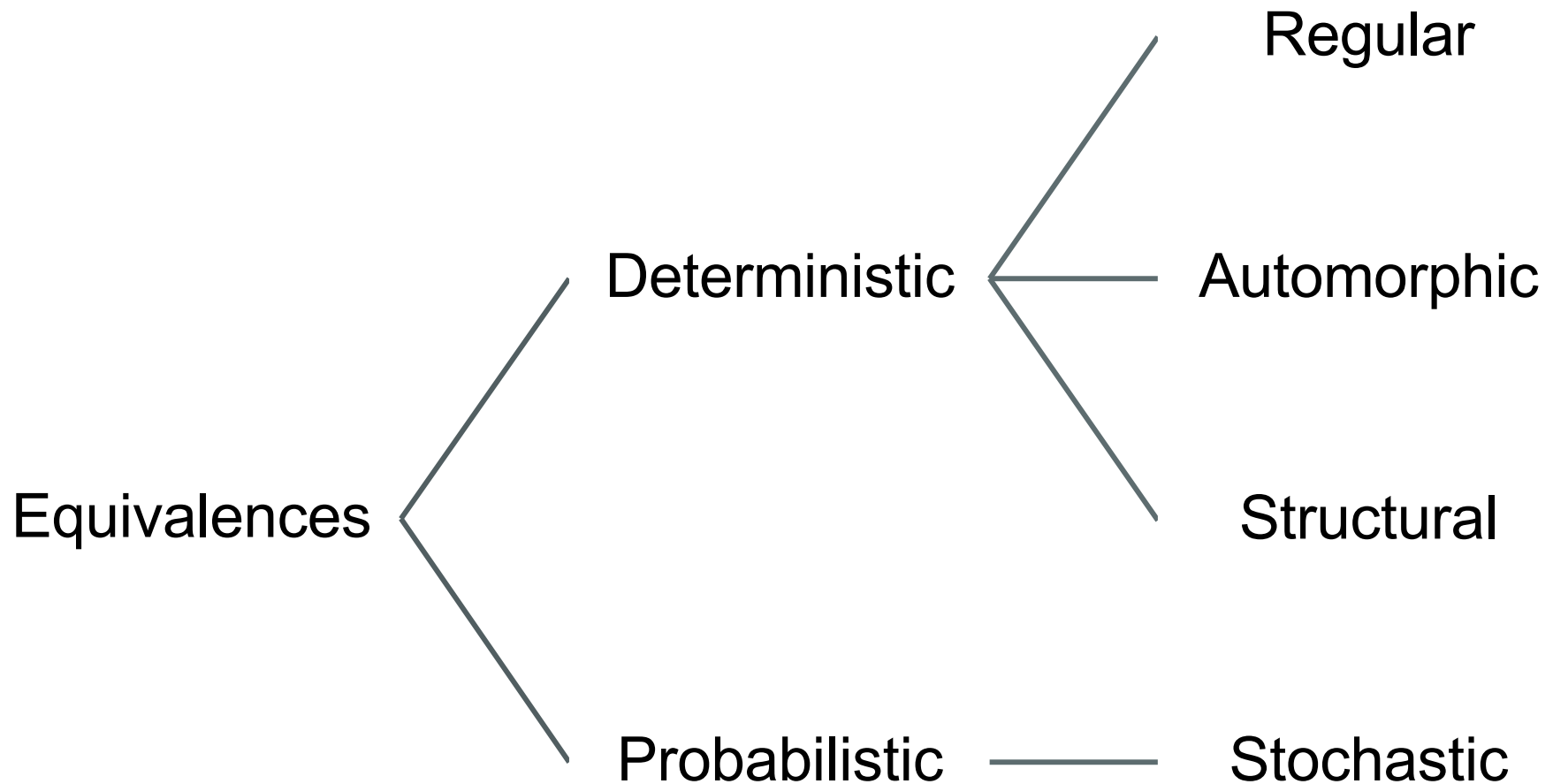
Roles find Regular Equivalences

Two nodes *u* and *v* are regularly equivalent if they are equally related to equivalent others.

Regular

[Everett & Borgatti, 1992]

Equivalences

Deterministic

Probabilistic

Automorphic

Structural

Stochastic
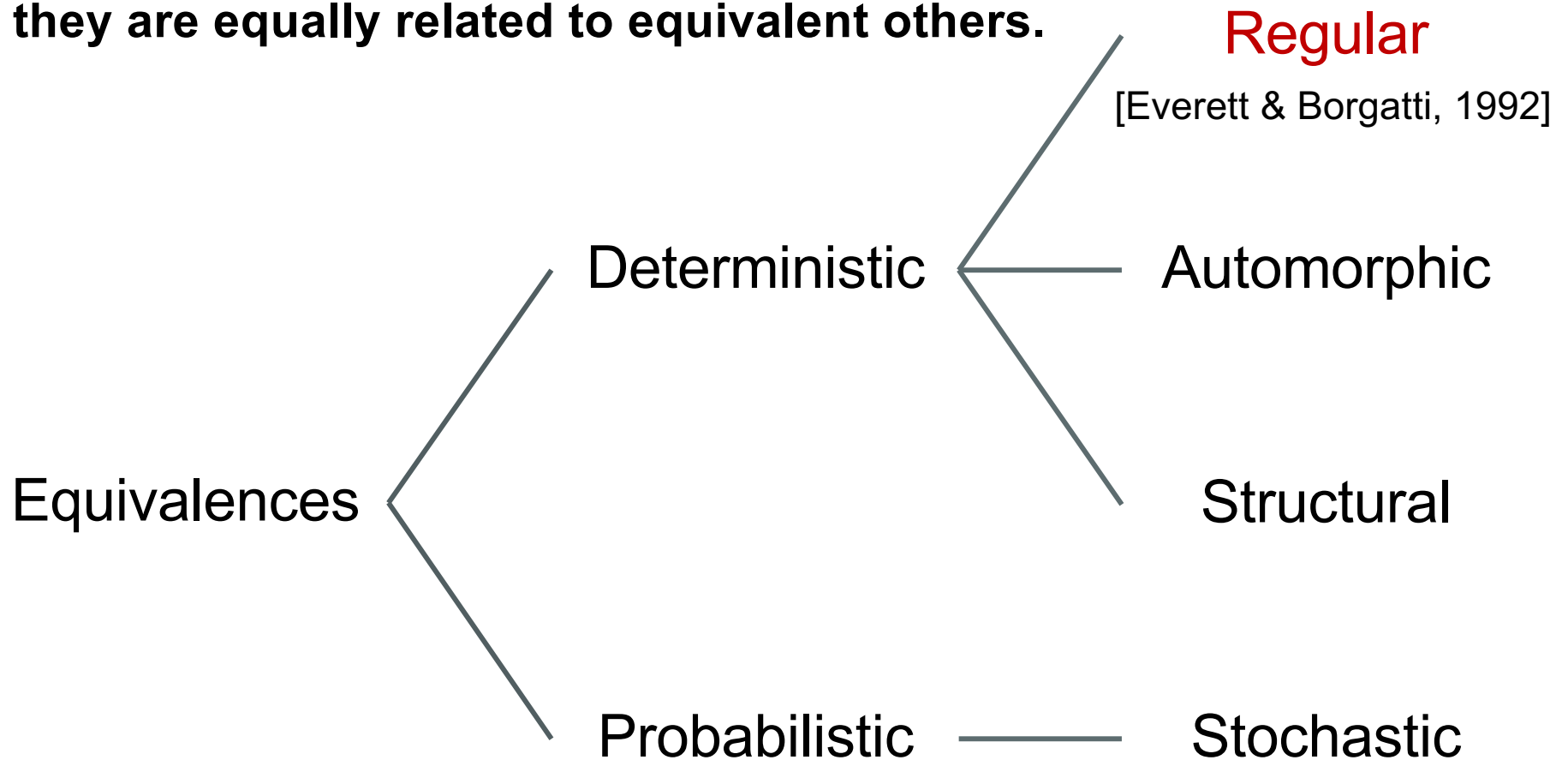
Finding Roles in a Network

Input

Node × Node
Matrix

n dim space

Finding Roles in a Network
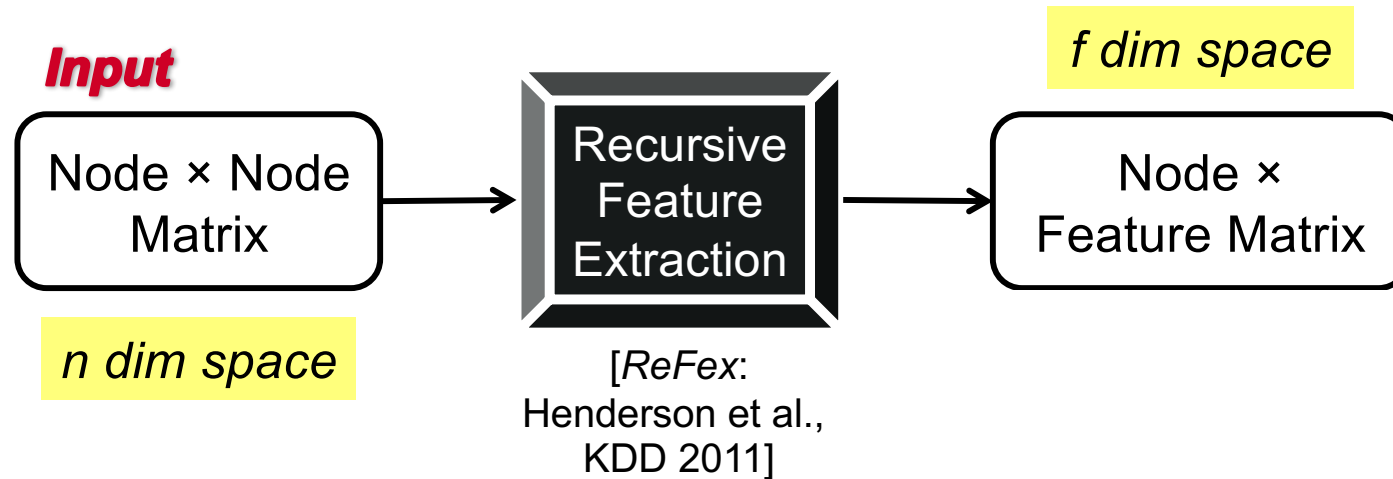
Input

Node × Node Matrix

n dim space

Recursive Feature Extraction

[*ReFex*: Henderson et al., KDD 2011]

f dim space

Node × Feature Matrix

ReFeX: Recursive Feature Extraction

- [Henderson *et al*., KDD 2011]

- Recursively combines node-based features with egonet-based features to output regional features
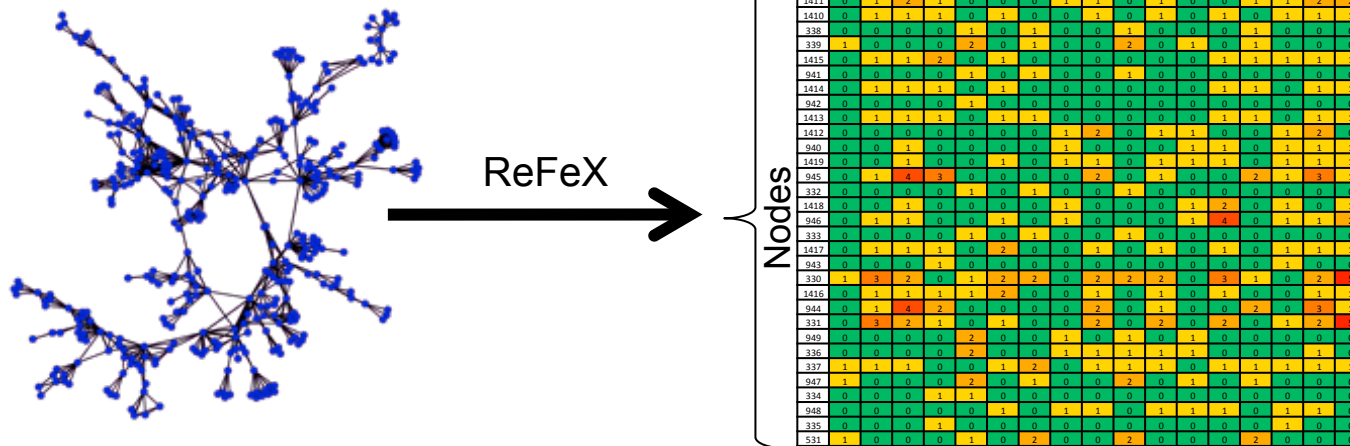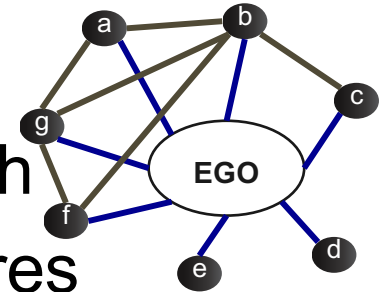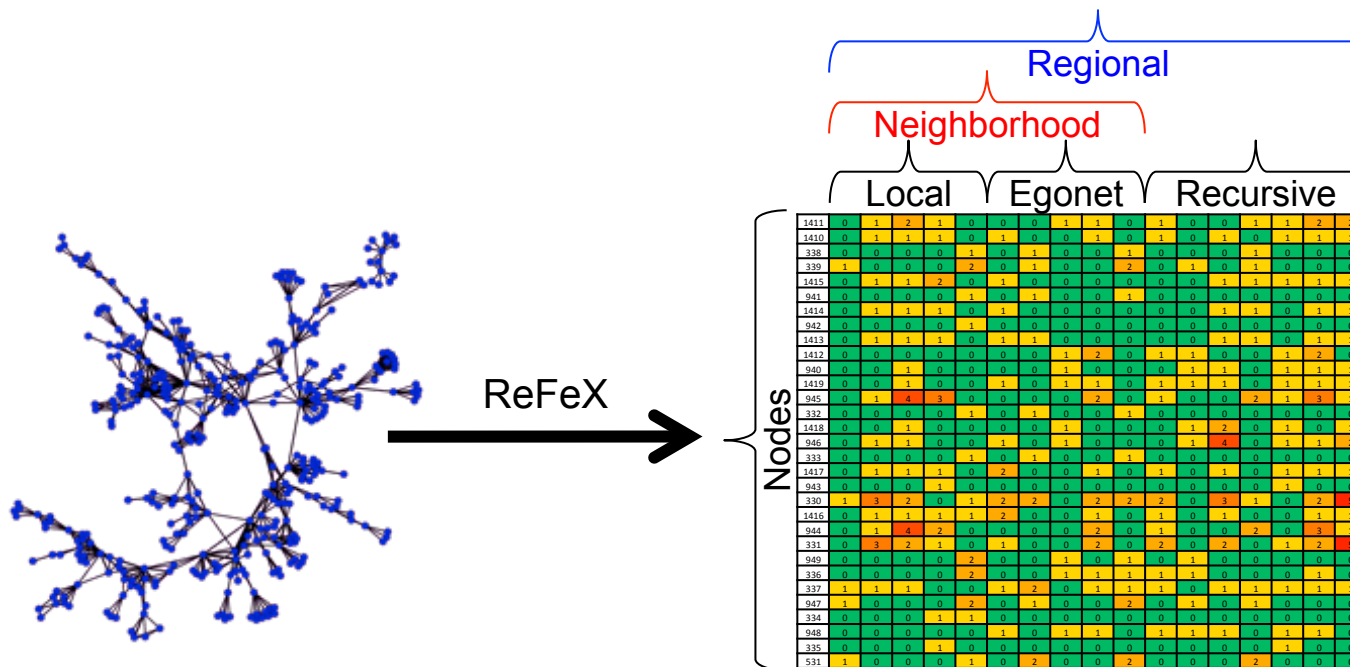
ReFeX: Recursive Feature Extraction

- [Henderson *et al*., KDD 2011]

- Recursively combines node-based features with egonet-based features to output regional features
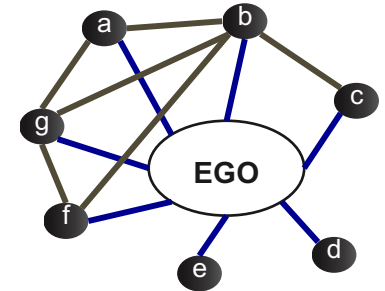
- Neighborhood features: What is your connectivity pattern?

- Recursive Features: To what kinds of nodes are you connected?

ReFeX: Structural Features

- 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

Degree

Source vs. Sink

Star vs. Cluster
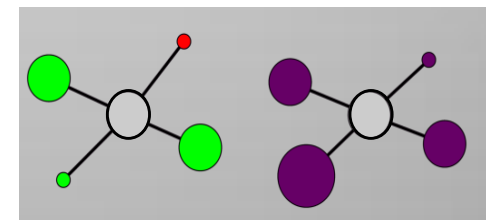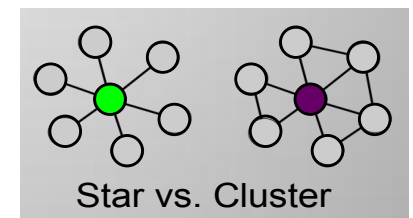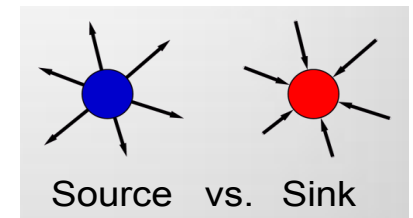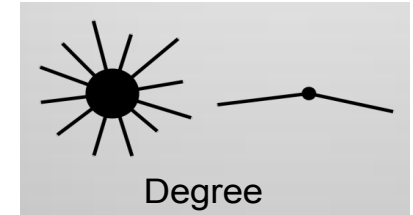
ReFeX: Structural Features

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

Degree

Source   vs.   Sink

Star vs. Cluster

ReFeX: Structural Features

Neighborhood

Regional

- **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

Degree

Source vs. Sink

Star vs. Cluster
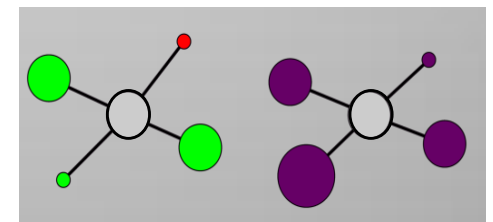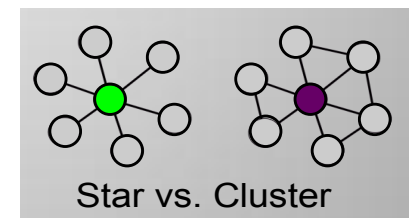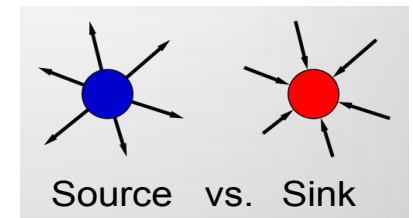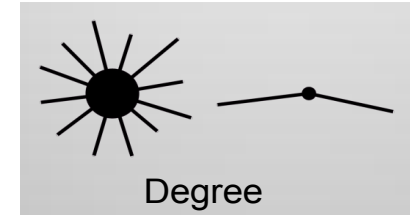
ReFeX (continued)

- Number of possible recursive features is infinite

ReFeX (continued)

- Number of possible recursive features is infinite

- ReFeX pruning

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
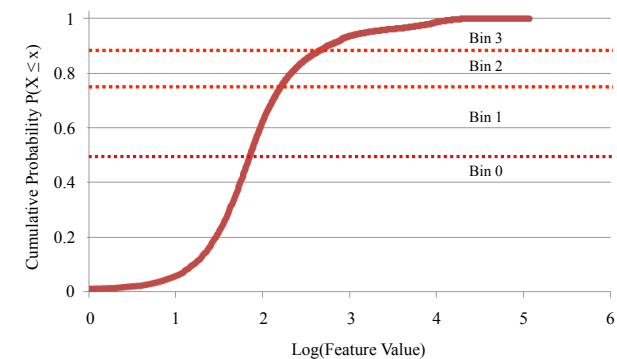
ReFeX (continued)

- 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

Finding Roles in a Network

Input

| Node × Node Matrix |

n dim space

Recursive Feature Extraction

[*ReFex*: Henderson et al., KDD 2011]

f dim space

| Node × Feature Matrix |

Finding Roles in a Network

Input

Node × Node Matrix

n dim space

Recursive Feature Extraction

[*ReFex*: Henderson et al., KDD 2011]

f dim space

Node × Feature Matrix

Role Extraction

[*Rolx*: Henderson et al., KDD 2012]

Node × Role Matrix

Role × Feature Matrix

r dim space

Output

Finding Roles in a Network

Input

Node × Node Matrix

n dim space

Recursive Feature Extraction

[*ReFex*: Henderson et al., KDD 2011]

f dim space

Node × Feature Matrix

Role Extraction

[*Rolx*: Henderson et al., KDD 2012]

n >> f >> r

Node × Role Matrix

Role × Feature Matrix

r dim space **Output**

Role Extraction: Feature Grouping

- Soft clustering in the structural feature space

  - Each node has a mixed-membership across roles
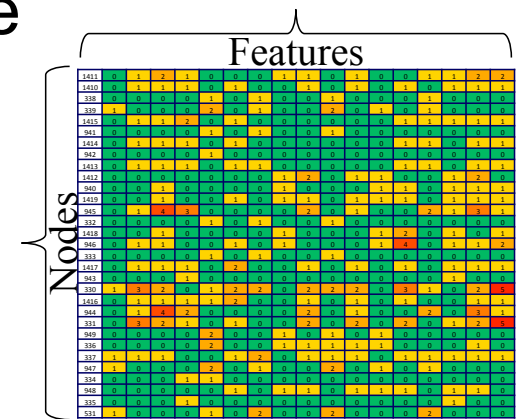
Role Extraction: Feature Grouping

- Soft clustering in the structural feature space

  - Each node has a mixed-membership across roles

- Generate a rank *r* approximation of *V ≈ GF*

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*

features   roles   features

nodes V ≈   nodes G ×   roles F
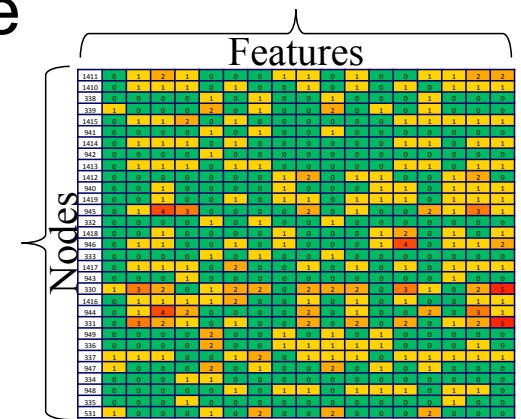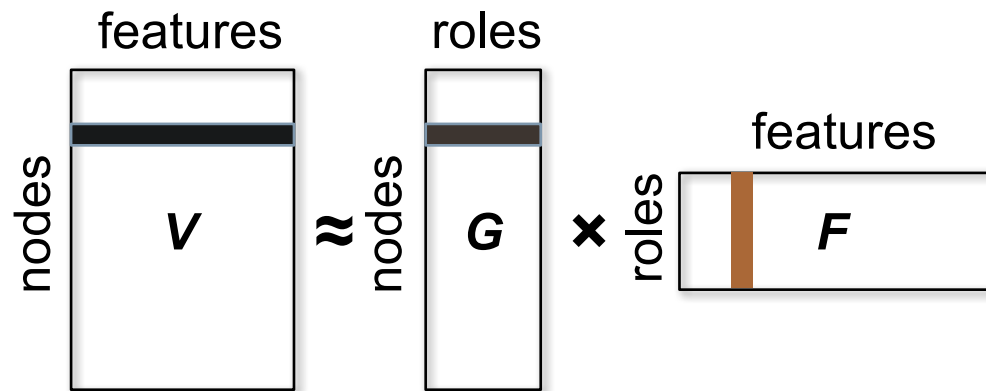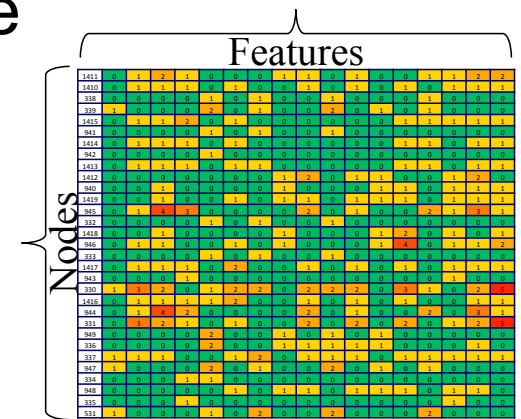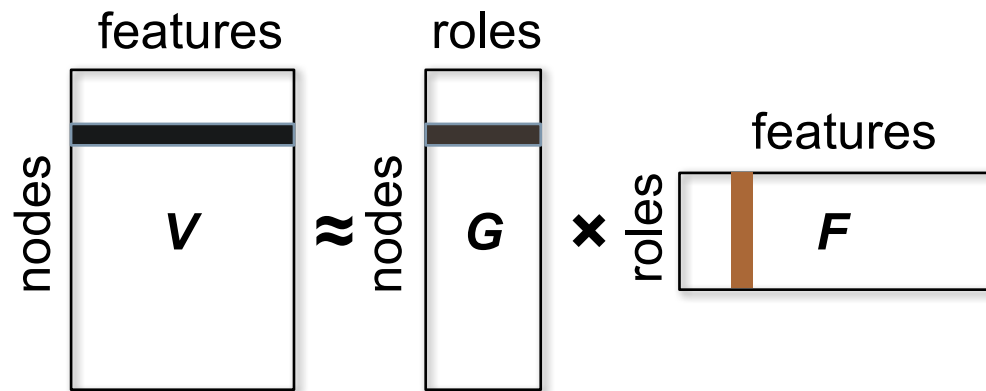
- ## RoIX uses NMF for feature grouping

  - ### Computationally efficient

  - ### Non-negative factors simplify interpretation of roles and memberships

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

Role Extraction: Model Selection

- Roles summarize behavior
  - Or, they compress the feature matrix, *V*

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 $L = M + E$

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

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

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

Finding Roles in a Network

Input

Node × Node Matrix

n dim space

Recursive Feature Extraction

[*ReFex*: Henderson et al., KDD 2011]

f dim space

Node × Feature Matrix

Role Extraction

[*Rolx*: Henderson et al., KDD 2012]

n >> f >> r

Node × Role Matrix

Role × Feature Matrix

r dim space   *Output*
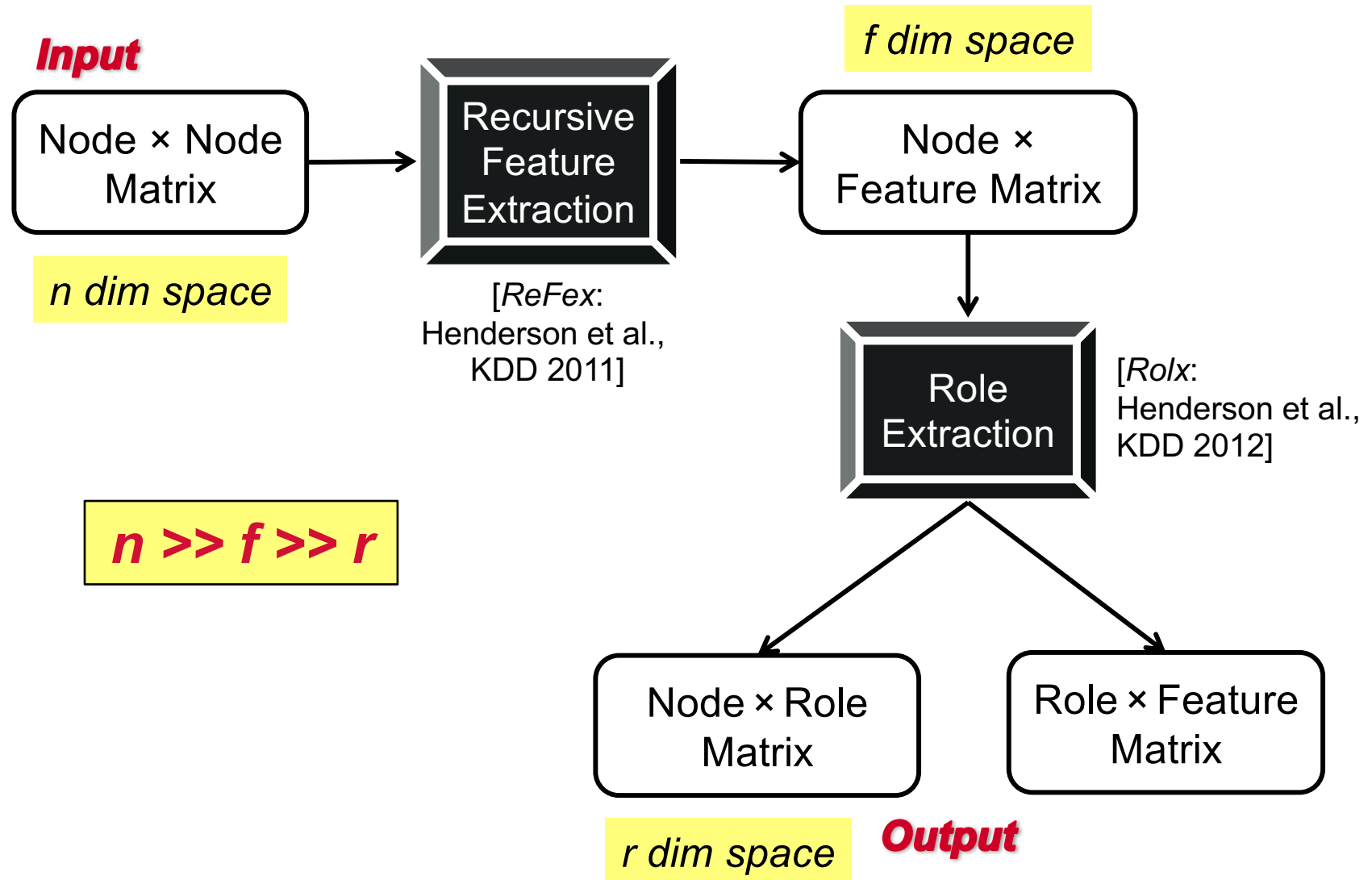
Automatically Discovered Roles
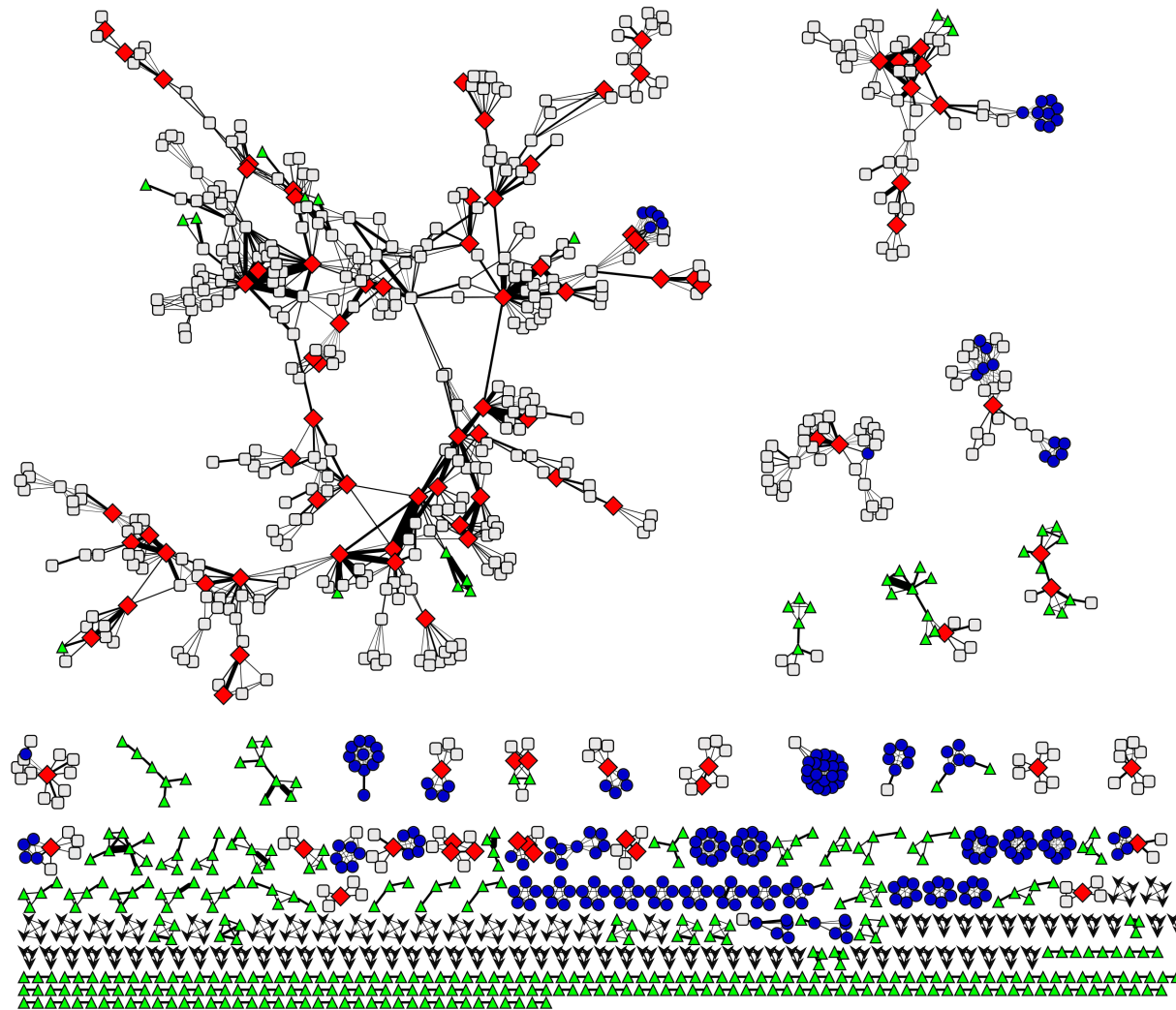
Network Science Co-authorship Graph
[Newman 2006]

Automatically Discovered Roles

◆ bridge
● cliquey
□ periphery
▲ isolated
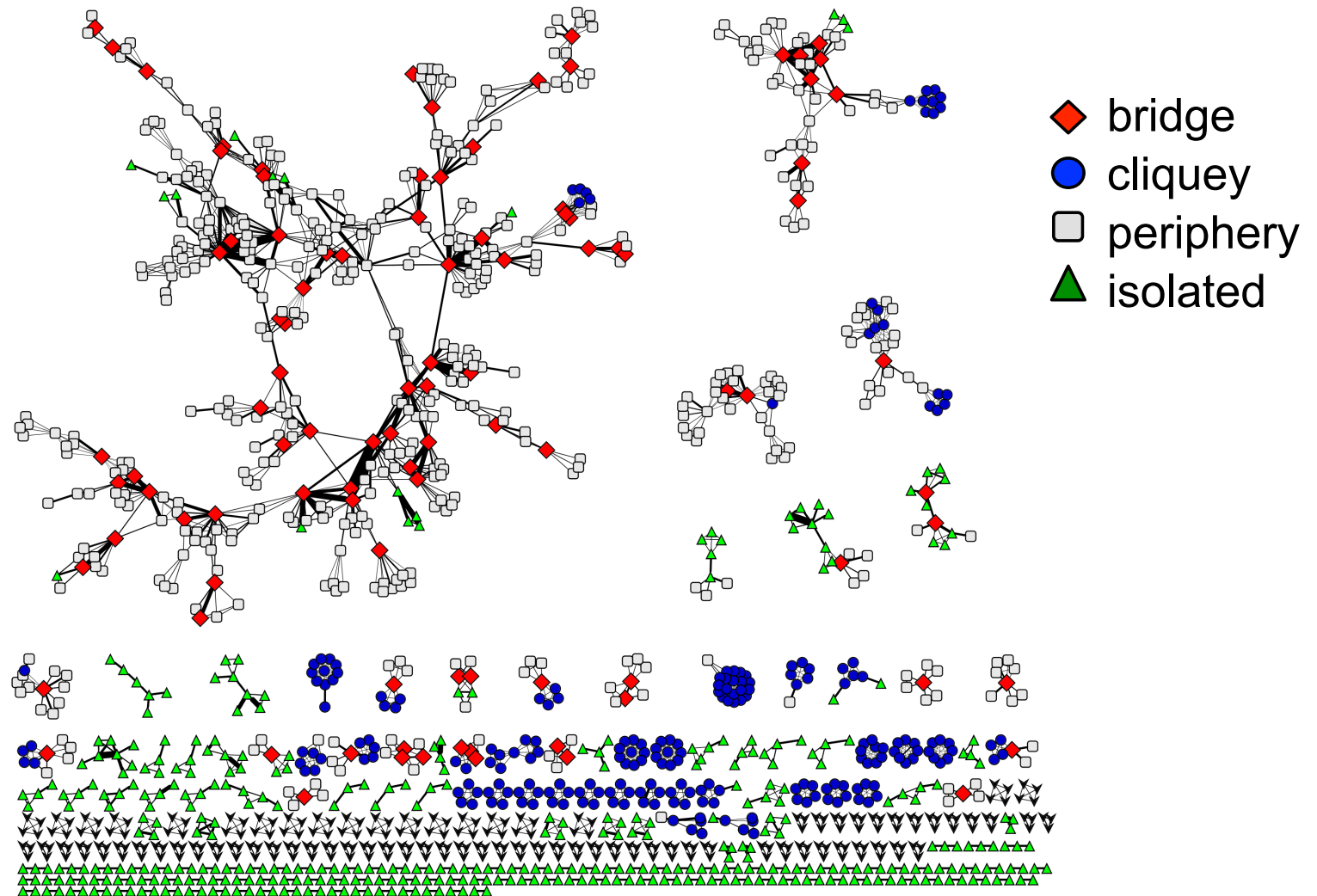
Network Science Co-authorship Graph
[Newman 2006]

Making Sense of Roles

Role-contribution to Node-measurement (y-axis, values 0 to 14)

Categories (x-axis): Degree, Avg. Weight, Clustering Coeff., Eccentricity, PageRank, GateKeeper, GateKeeper-Local, Pivot, Structural Hole

Node Sense: GE≈M

Legend: Role 1 (blue), Role 2 (red), Role 3 (gray), Role 4 (green), Default (dashed orange line)

cliquey — *bridge* — *periphery* — *isolated*

Making Sense of Roles

Node Sense: $GE \approx M$ **Neighbor Sense:** $GQ \approx N$

Role 1   Role 2   Role 3   Role 4   Default

cliquey   *bridge*   periphery   *isolated*

Making Sense of Roles

Topological measures & role homophily help interpret roles.

Chart: Role-contribution to Node-measurement (y-axis, scale 0 to 14) across topological measures. X-axis measures under *Node Sense: GE≈M*: Degree, Avg. Weight, Clustering Coeff., Eccentricity, PageRank, GateKeeper, GateKeeper-Local, Pivot, Structural Hole. Under *Neighbor Sense: GQ≈N*: Role 1, Role 2, Role 3, Role 4 (Role Affinities).

Legend:
- Role 1 (blue) — *cliquey*
- Role 2 (red) — *bridge*
- Role 3 (gray) — periphery
- Role 4 (green) — *isolated*
- Default (orange dashed line)

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 comparison	Determine network compatibility for knowledge transfer
Exploration in role space	Exploratory analysis of network data in the role space
…	…

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 comparison	Determine network compatibility for knowledge transfer
Exploration in role space	Exploratory analysis of network data in the role space
…	…

O'REILLY®

Search
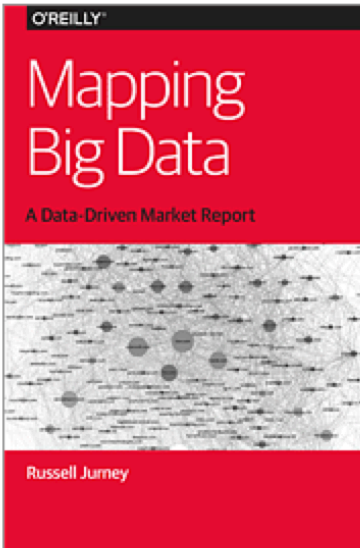
Mapping Big Data
A Data-Driven Market Report

By Russell Jurney

Publisher: O'Reilly
Released: September 2015

Description

To discover the shape and structure of the big data market, the San Francisco-based startup Relato took a unique approach to market research and created the first fully data-driven market report. Company CEO Russell Jurney and his team collected and analyzed raw data from a variety of sources to reveal a boatload of business insights about the big data space. This exceptional report is now available for free download.

Using data analytic techniques such as social network analysis (SNA), Relato exposed the vast and complex partnership network that exists among tens of thousands of unique big data vendors. The dataset Relato collected is centered around Cloudera, Hortonworks, and MapR, the major platform vendors of Hadoop, the primary force behind this market.

From this snowball sample, a 2-hop network, the Relato team was able to answer several questions, including:

- Who are the major players in the big data market?
- Which is the leading Hadoop vendor?
- What sectors are included in this market and how do they relate?
- Which among the thousands of partnerships are most important?
- Who's doing business with whom?

Metrics used in this report are also visible in Relato's interactive web application, via a link in the report, which walks you through the insights step-by-step.

Russell Jurney is CEO of Relato, a San Francisco area startup that maps markets to drive sales and marketing. He is the author of Agile Data Science *and co-author of* Big Data for Chimps *(both O'Reilly). In addition, Russell is an Apache Committer on the Incubating DataFu project. Russell is a full stack engineer.*

Big data business partnerships

Russell Jurney
http://relato.io
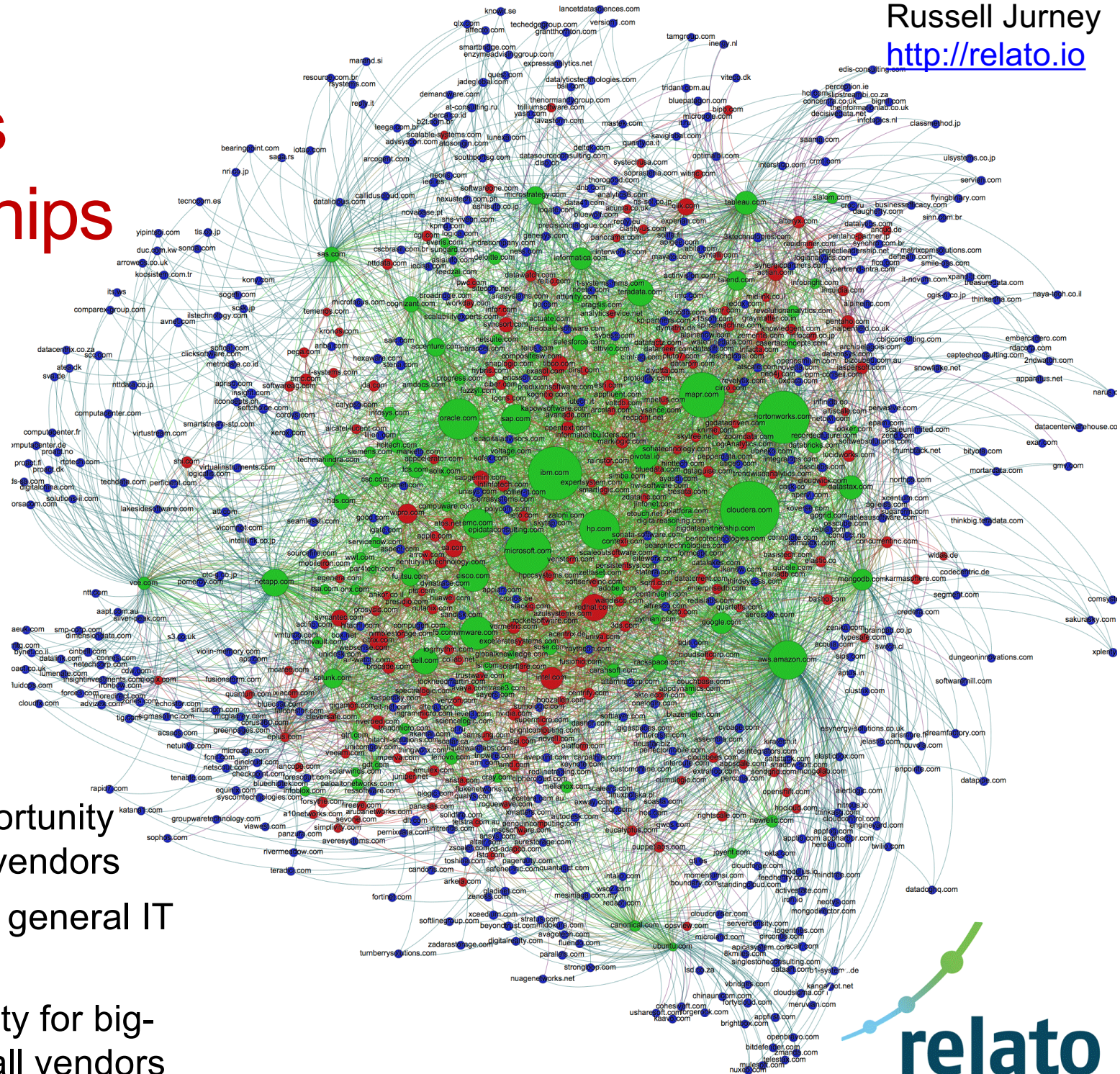
- **Green**: equal opportunity bridges; big-data vendors

- **Red**: middle-men; general IT vendors

- **Blue**: Strong affinity for big-data vendors; small vendors

relato

Big-data business-partnerships
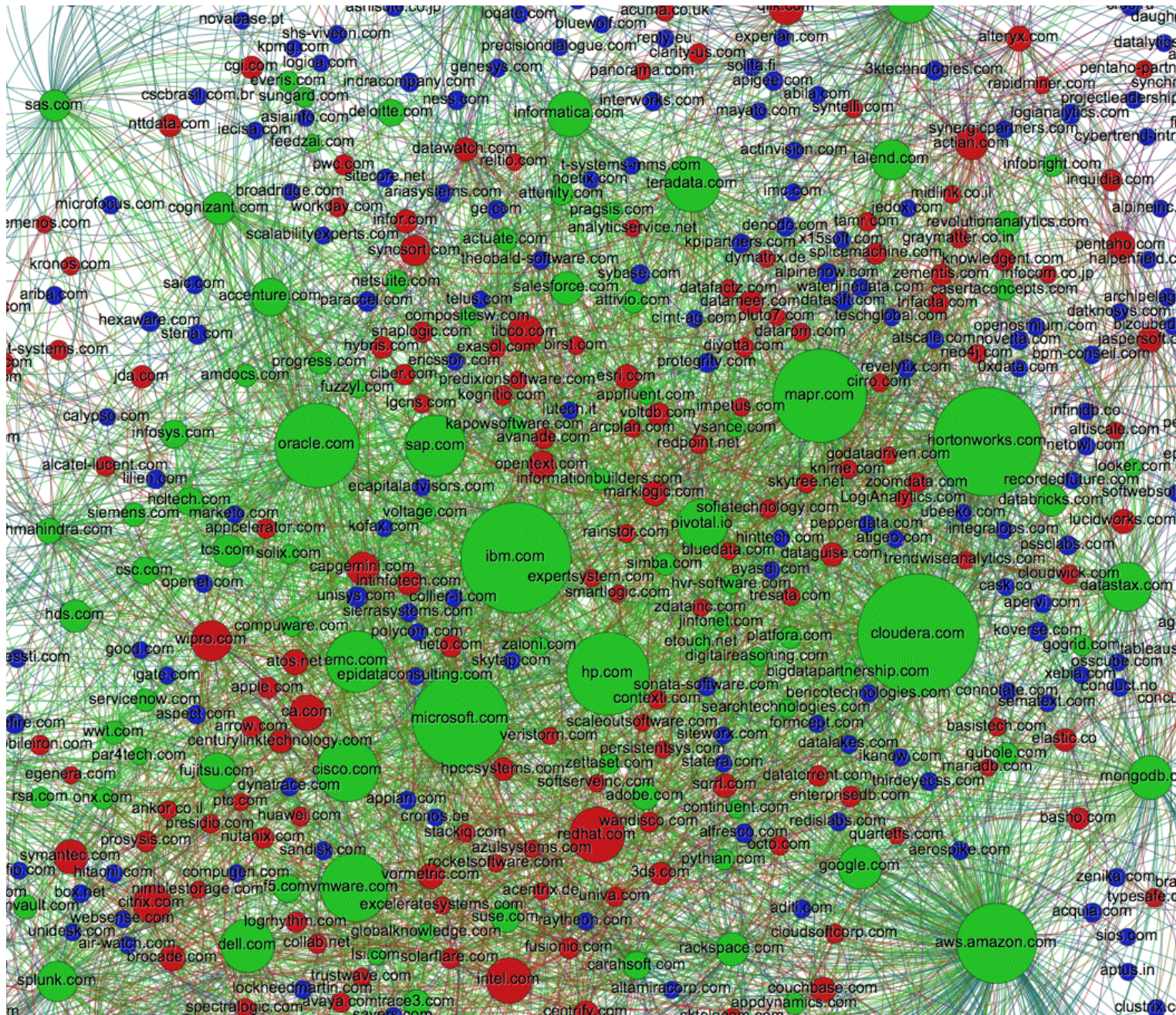
Louvain Clustering After Removing Small Vendors (Blue Role)

Russell Jurney
http://relato.io

Legend:
- Analytics Software
- Cloud Computing
- Enterprise Software
- New Data Platforms
- Old Data Platforms
- Servers

relato
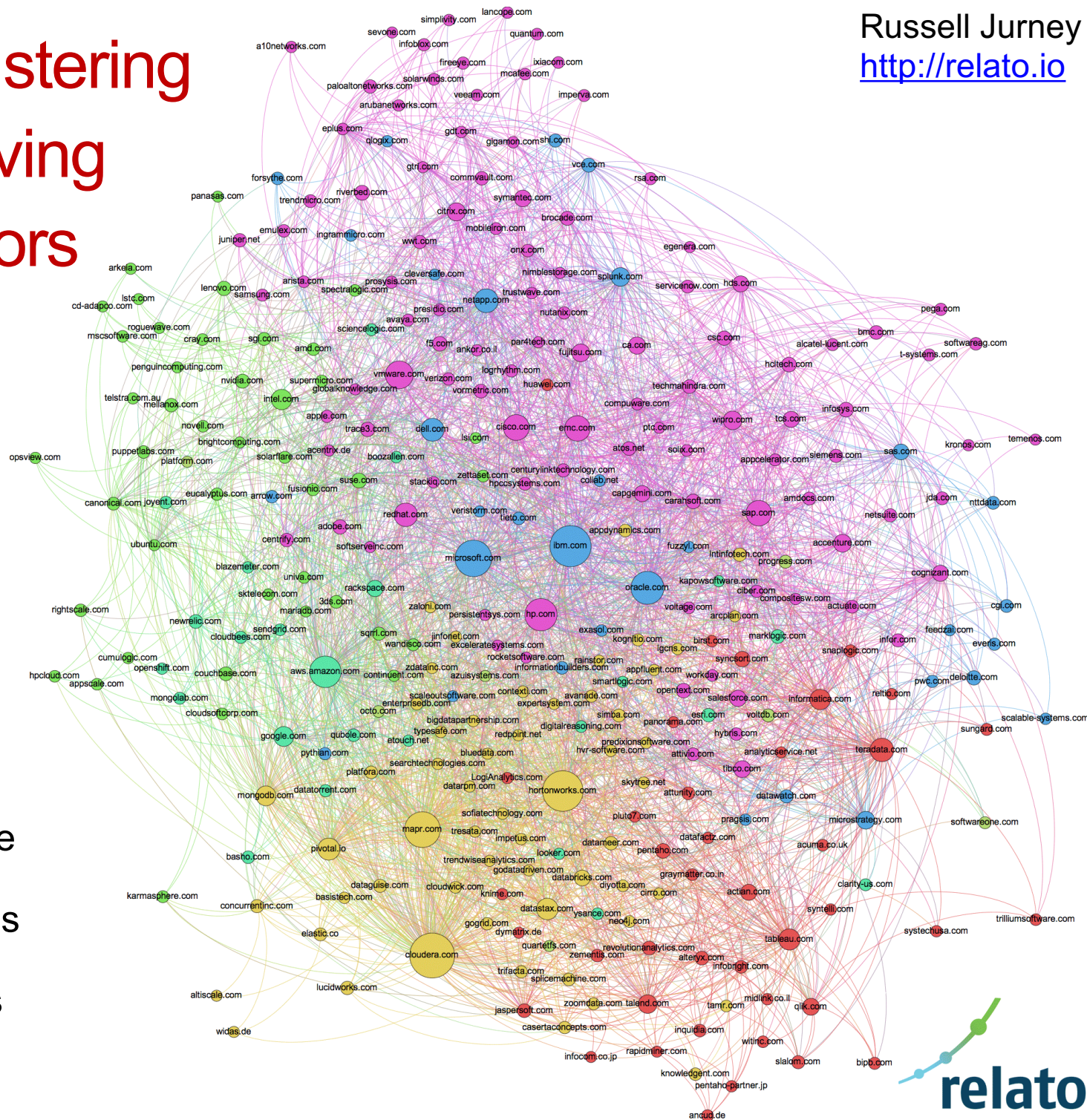
An Interactive Market Map of the Big Data Space

- http://demo.relato.io/oreilly and http://demo.relato.io/public

Role Transfer = RolX + SL

Role Transfer = RolX + SL

External Network

(1) RolX

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

Role Memberships

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

Role Definitions

Role Transfer = RolX + SL

External Network

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

Role Memberships

(1) RolX

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

Role Definitions

(2) Learning

	class
n1	🟢
n2	🔵
n3	🔴
⋮	

Classifier

P(class | role membership)

Role Transfer = RolX + SL

External Network

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

Role Memberships

(1) RolX

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

Role Definitions

	class
n1	🟢
n2	🔵
n3	🔴
⋮	

(2) Learning

Classifier

P(class | role membership)

Target Network

(3) Role Assignment

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

Role Memberships

Role Transfer = RolX + SL

External Network

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

Role Memberships

(1) RolX

(2) Learning

	class
n1	🟢
n2	🔵
n3	🔴
⋮	

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

Role Definitions

Classifier

$P(\text{class} \mid \text{role membership})$

(3) Role Assignment

Target Network

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

Role Memberships

(4) Inference

Target Network

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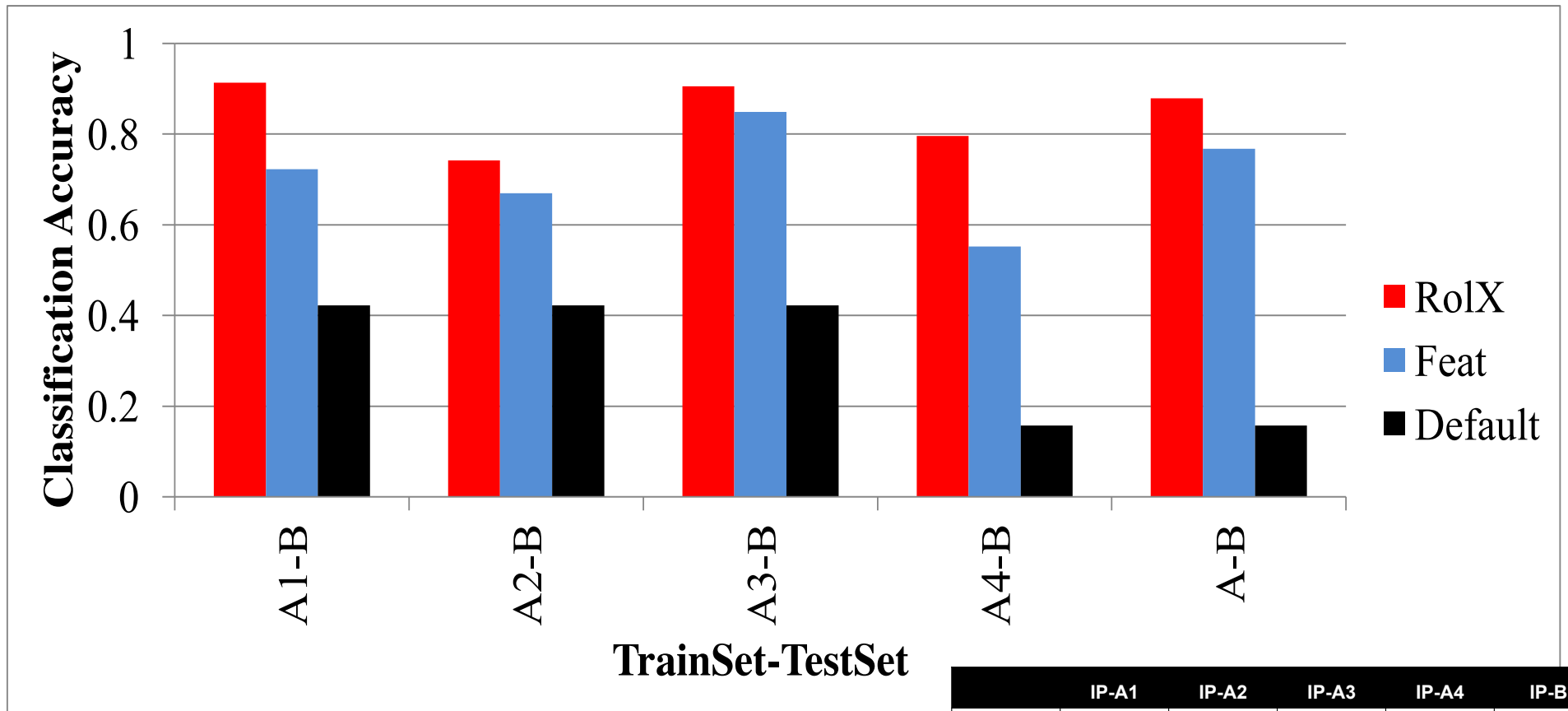
	IP-A1	IP-A2	IP-A3	IP-A4	IP-B
# Nodes	81,450	57,415	154,103	206,704	181,267
% labeled	36.7%	28.1%	20.1%	32.9%	15.3%
# Links	968,138	432,797	1,266,341	1,756,082	1,945,215
(# unique)	206,112	137,822	358,851	465,869	397,925
Class Distribu-tion					

Web DNS P2P

RolX Model Selection

9 roles automatically discovered by RolX
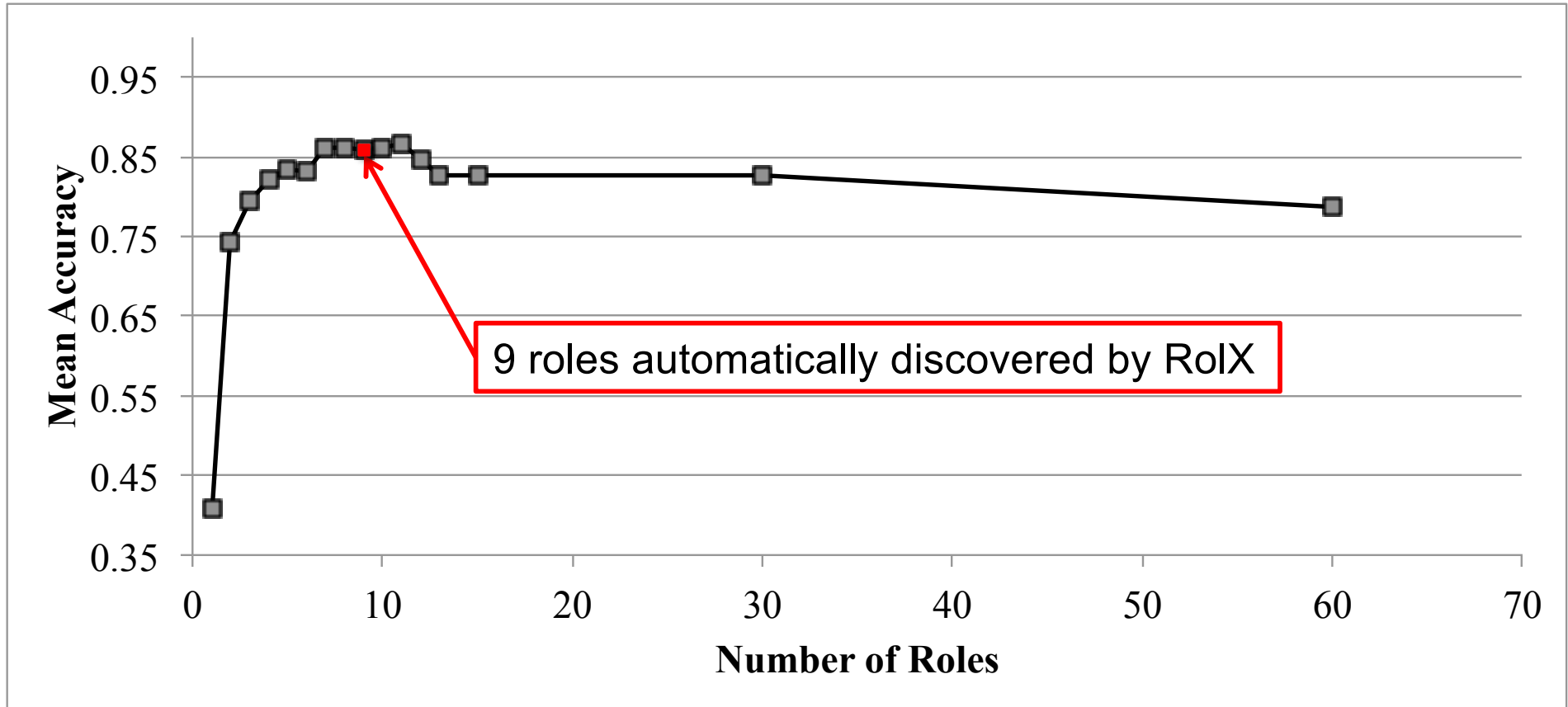
RolX selects high accuracy model sizes

RolX Model Selection (continued)
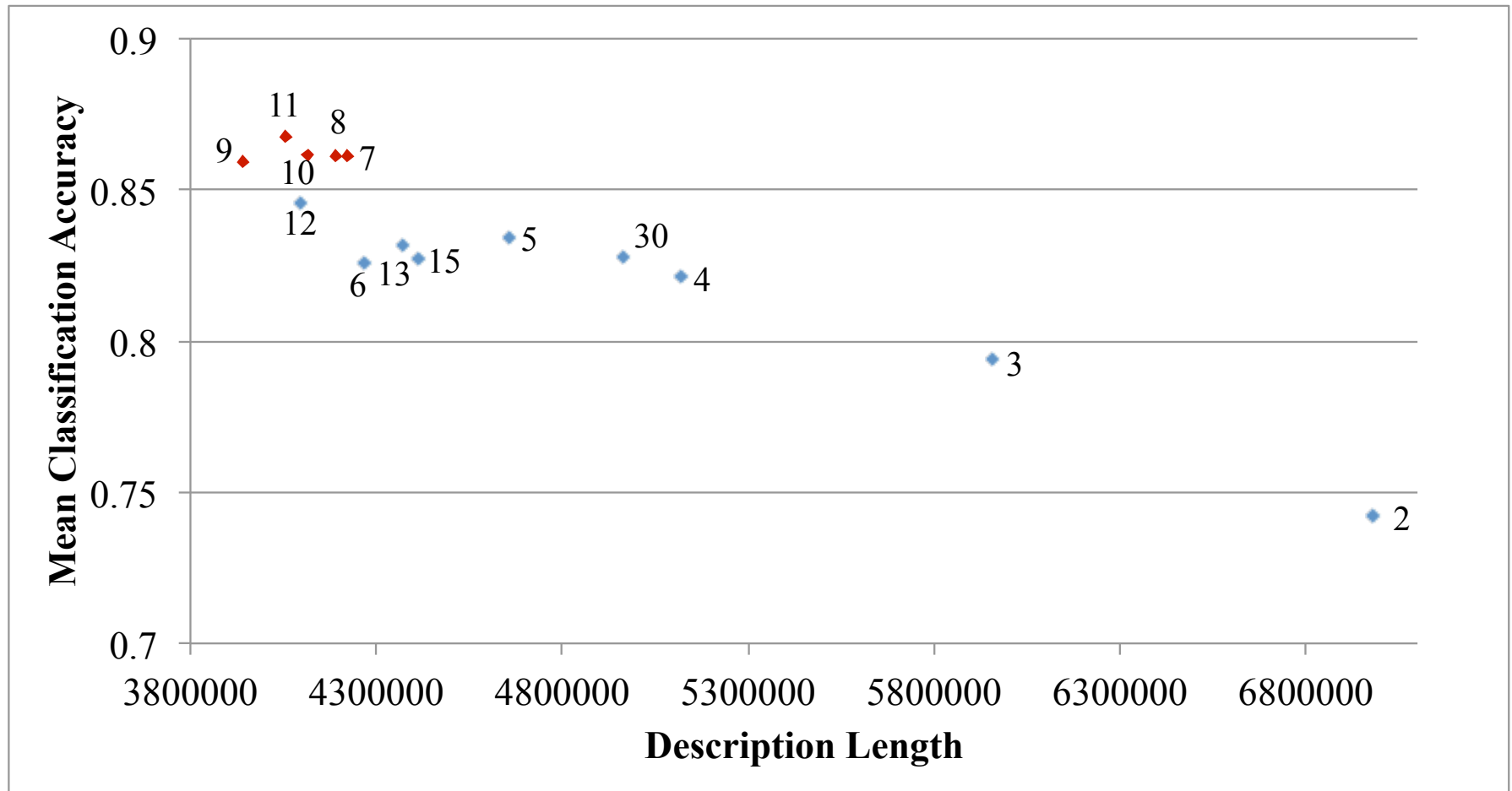
Classification accuracy is highest when RolX selection criterion is minimized

# 2nd Generation Algorithms for Role Discovery

- *GLRD*: guided learning for role discovery

  - [Gilpin et al., KDD 2013]

- *DBMM*: dynamic behavioral mixed-membership model

  - [Rossi et al., WSDM 2013]

- *RC-Joint*: simultaneous detection of communities and roles

  - [Ruan & Parthasarathy, COSN 2014]

- Motif-Role-Fingerprints

  - [McDonnell et al., PLoS ONE 9(12), 2014]

- Dynamic inference of social roles in information cascades

  - [Choobdar et al., DMKD 29(5), 2015]

- *MRD*: multi-relational role discovery

  - [Gilpin et al., ArXiv 2016]

- *DERM:* dynamic edge role mixed-membership model

  - [Ahmed et al., ArXiv 2016]

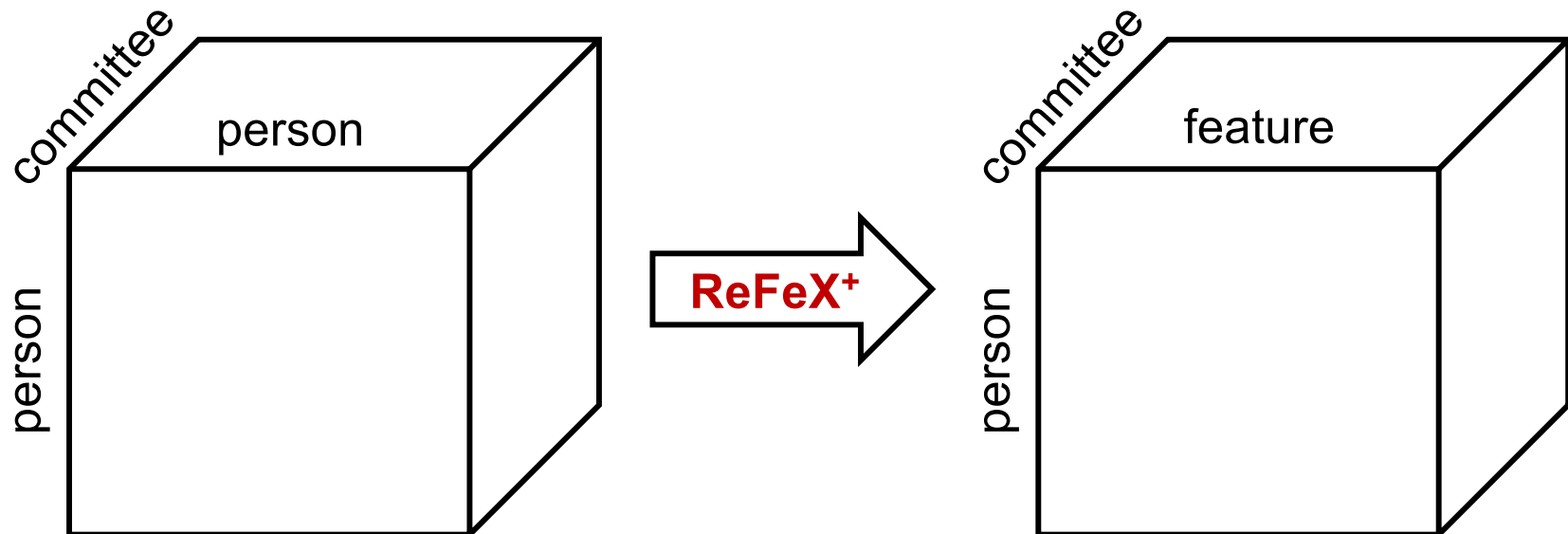- A combinatorial approach to role discovery

  - [Arockiasamy et al., ICDM 2016]

2nd Generation Algorithms for Role Discovery

- *GLRD*: guided learning for role discovery
  - [Gilpin et al., KDD 2013]
- *DBMM*: dynamic behavioral mixed-membership model
  - [Rossi et al., WSDM 2013]
- *RC-Joint*: simultaneous detection of communities and roles
  - [Ruan & Parthasarathy, COSN 2014]
- Motif-Role-Fingerprints
  - [McDonnell et al., PLoS ONE 9(12), 2014]

- Dynamic inference of social roles in information cascades
  - [Choobdar et al., DMKD 29(5), 2015]
- *MRD*: multi-relational role discovery
  - [Gilpin et al., ArXiv 2016]
- *DERM:* dynamic edge role mixed-membership model
  - [Ahmed et al., ArXiv 2016]
- A combinatorial approach to role discovery
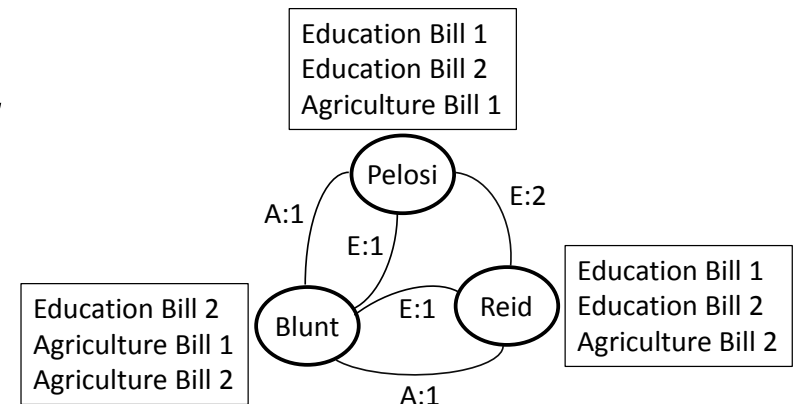 - [Arockiasamy et al., ICDM 2016]

Multi-relational Role Discovery (MRD)

- Moving beyond simple networks

- Suppose you have a multi-relational networks

- Example: Congressional co-sponsorship data

[Gilpin et al., ArXiv 2016]
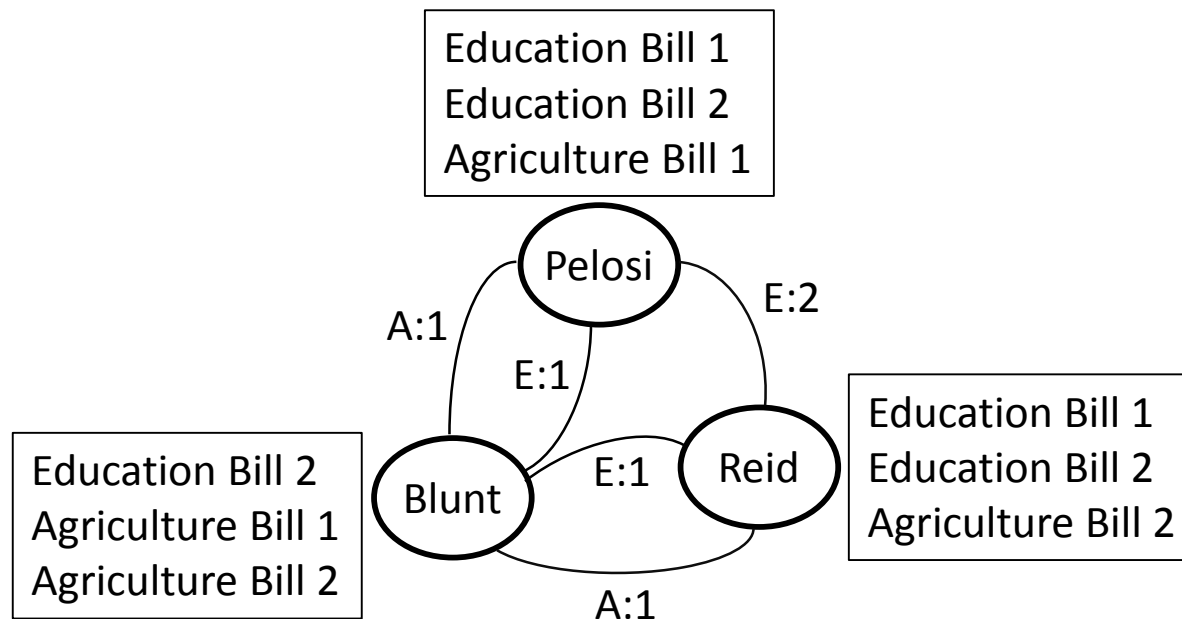
No longer have an adjacency matrix

- We have a `person × person × committee` tensor

- Entry at (i , j, k) indicates
how often congress-person i and j
co-sponsored a bill that was sent
to committee k for a particular
congressional committee

Education Bill 1
Education Bill 2
Agriculture Bill 1

Pelosi

A:1   E:2

E:1

Education Bill 2
Agriculture Bill 1
Agriculture Bill 2

Blunt   E:1   Reid

Education Bill 1
Education Bill 2
Agriculture Bill 2

A:1

committee

person

person

ReFeX⁺

committee

feature

person

Multi-relational Role Discovery (MRD)

- *Multi-relational Role Discovery* (MRD)

  - No orthogonality constraint on factors

 - Nonnegative Tucker decomposition

  - Alternating least squares

[Gilpin et al., ArXiv 2016]
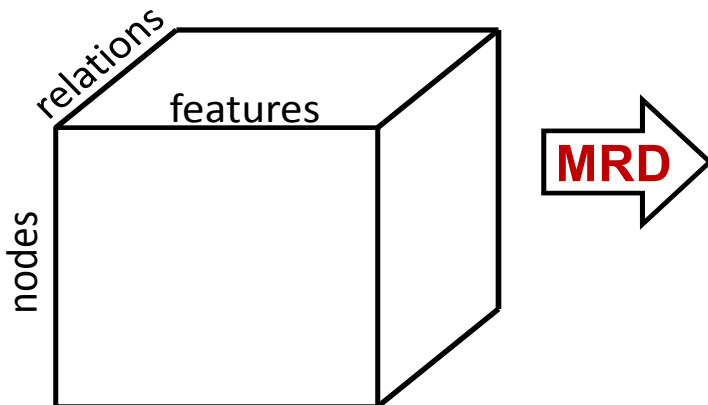
Multi-relational Role Discovery (MRD)

- *Multi-relational Role Discovery* (MRD)

  - No orthogonality constraint on factors

  - Nonnegative Tucker decomposition

  - Alternating least squares

$$\operatorname*{argmin}_{\mathbf{G},\mathbf{F},\mathbf{R},\mathcal{H}} \quad ||\mathcal{V} - \sum_i \sum_j \sum_k h_{ijk} * \mathbf{g_k} \circ \mathbf{f_k} \circ \mathbf{r_k}||_{Fro}$$

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

$$g_i(\mathcal{H}) \leq d_{\mathcal{H}_i}, i = 1 \dots t_{\mathcal{H}}$$

$$where \ g_i \ is \ a \ convex \ function$$
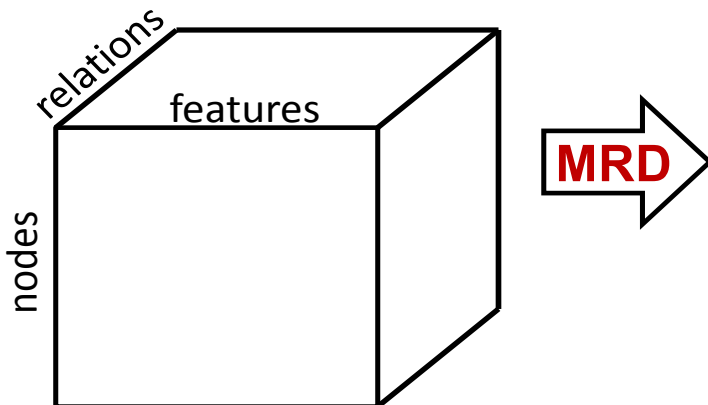
[Gilpin et al., ArXiv 2016]

Multi-relational Role Discovery (MRD)

- *Multi-relational Role Discovery (MRD)*
  - No orthogonality constraint on factors
  - Nonnegative Tucker decomposition
  - Alternating least squares

- The factor matrices are:
  - groups of features (*role definitions*)
  - groups of entities (*groups*)
  - groups of relations (*topics*)
- Tucker core
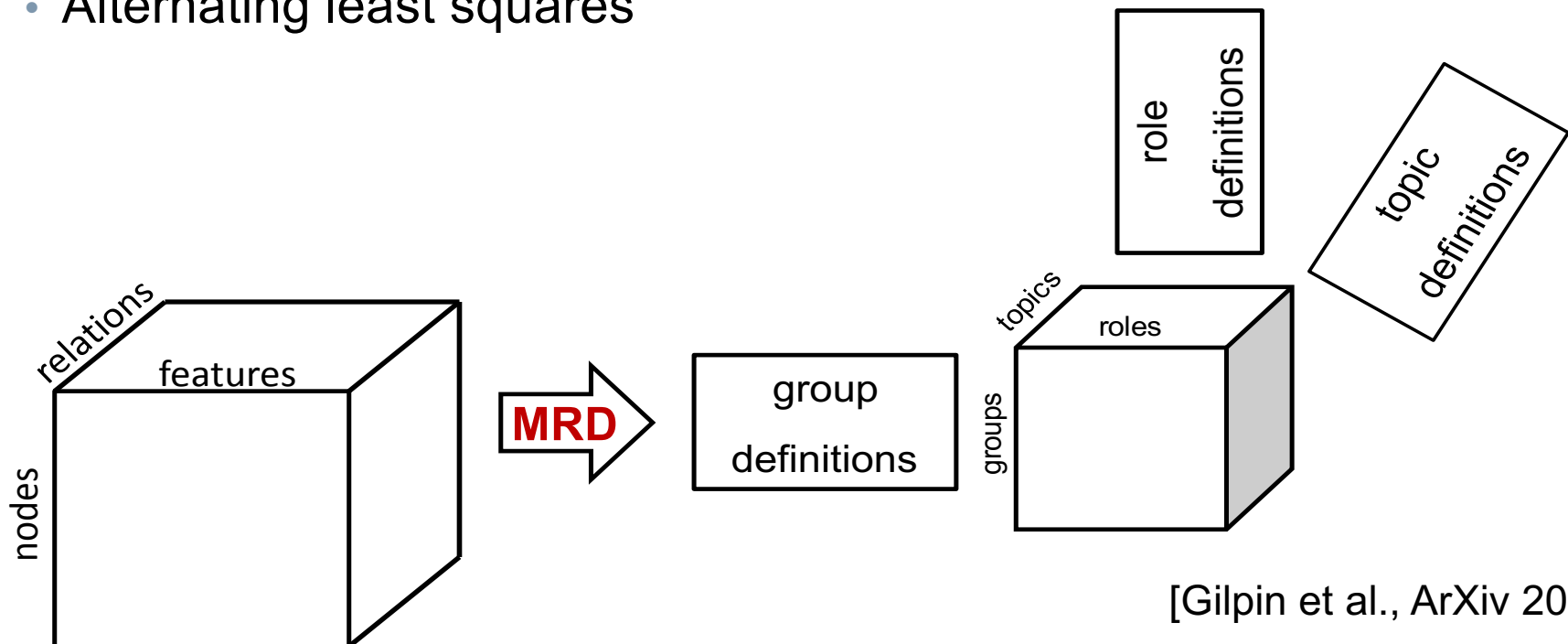
[Gilpin et al., ArXiv 2016]

Experiments

- Data from U.S. House of Representatives

- Bill co-sponsorship data from 1979 (the start of the 96th Congress) to 2009 (the end of the 110th Congress)

- 15 committees, for which there were legislation in each congress from 96th to 110th

- 110th Congress (from 2007-09)

  - 453 representatives & 10,613 bills

  - Average degree in aggregated graph = 8.37

  - Median value of average degree across committee co-sponsorship graphs = 0.48

| |
|---|
| Sci & Tech |
| Judiciary |
| Ways & Means |
| VA |
| Small Business |
| Budget |
| Oversight & Gov't Reform |
| Agriculture |
| Appropriations |
| Rules |
| Natural Resources |
| Financial Services |
| Education & Labor |
| Transportation & Infrastructure |
| Energy & Commerce |

Multi-relational Role Discovery (MRD)

[Gilpin et al., ArXiv 2016]
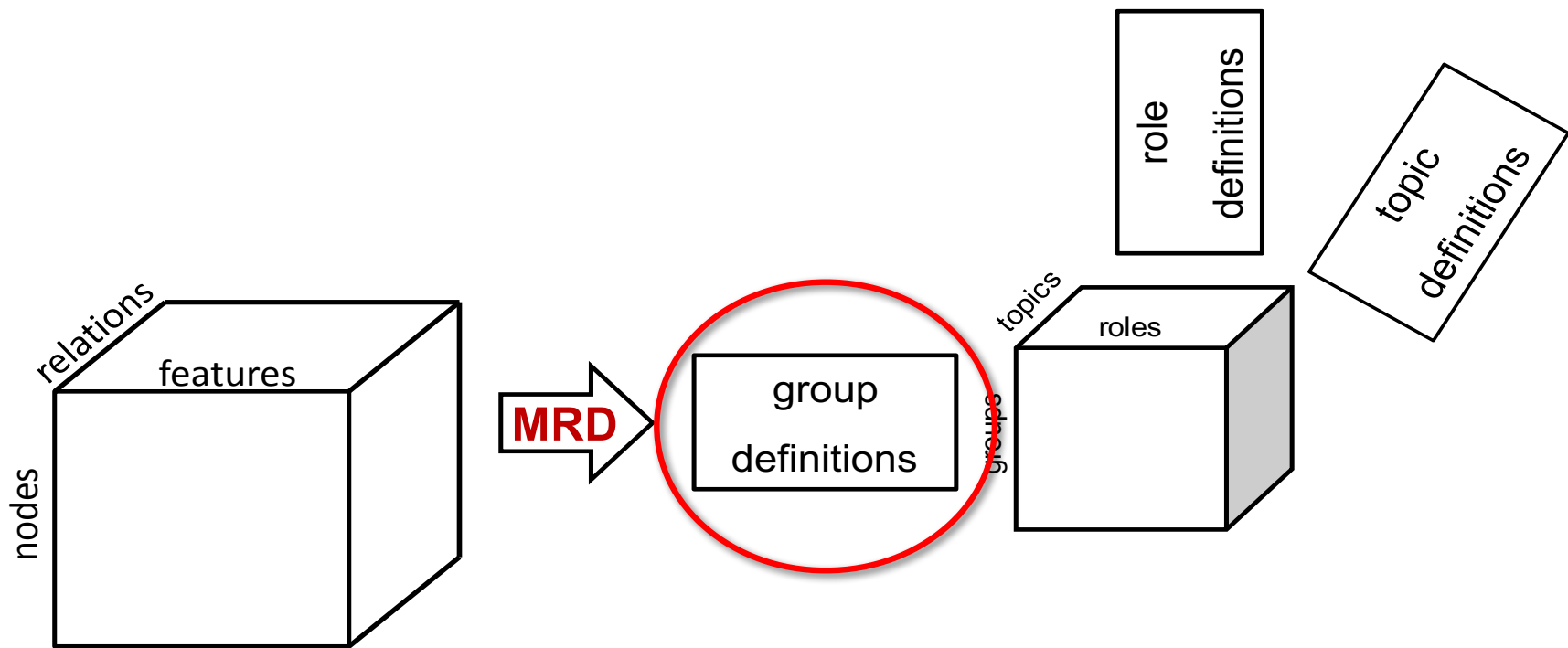
Groups of representatives

Group Members 1
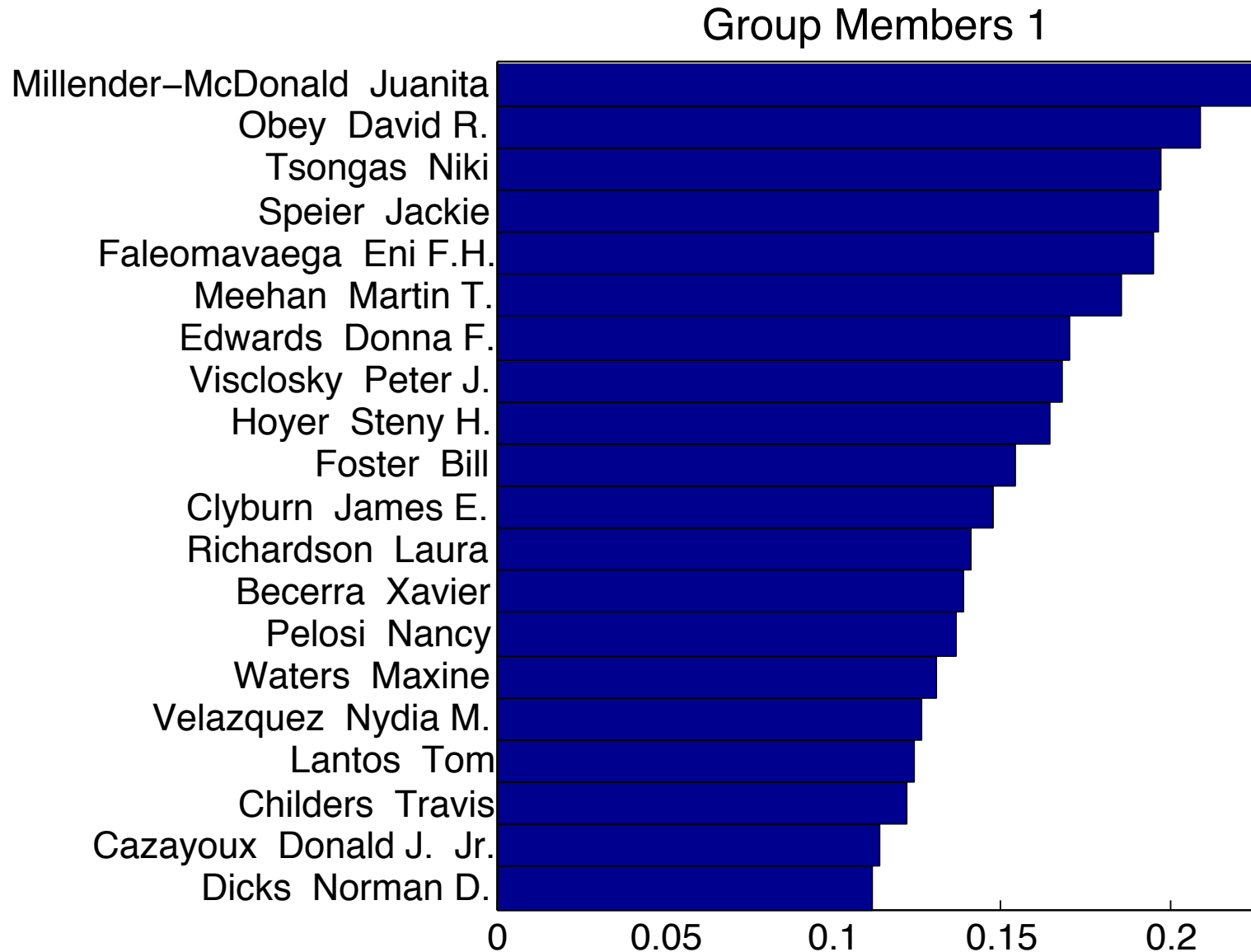
Group 1 of representatives

Name	Party	Exp
Millender-McDonald	D	11
Obey, David	D	38
Tsongas, Niki	D	0
Speier, Jackie	D	0
Faleomavaega, Eni	D	18
Meehan, Martin	D	14
Edwards, Donna	D	0
Visclosky, Peter	D	22
Hoyer, Steny	D	26
Foster, Bill	D	0
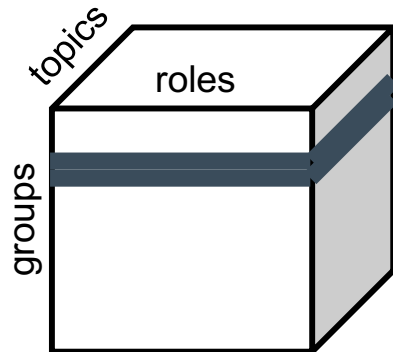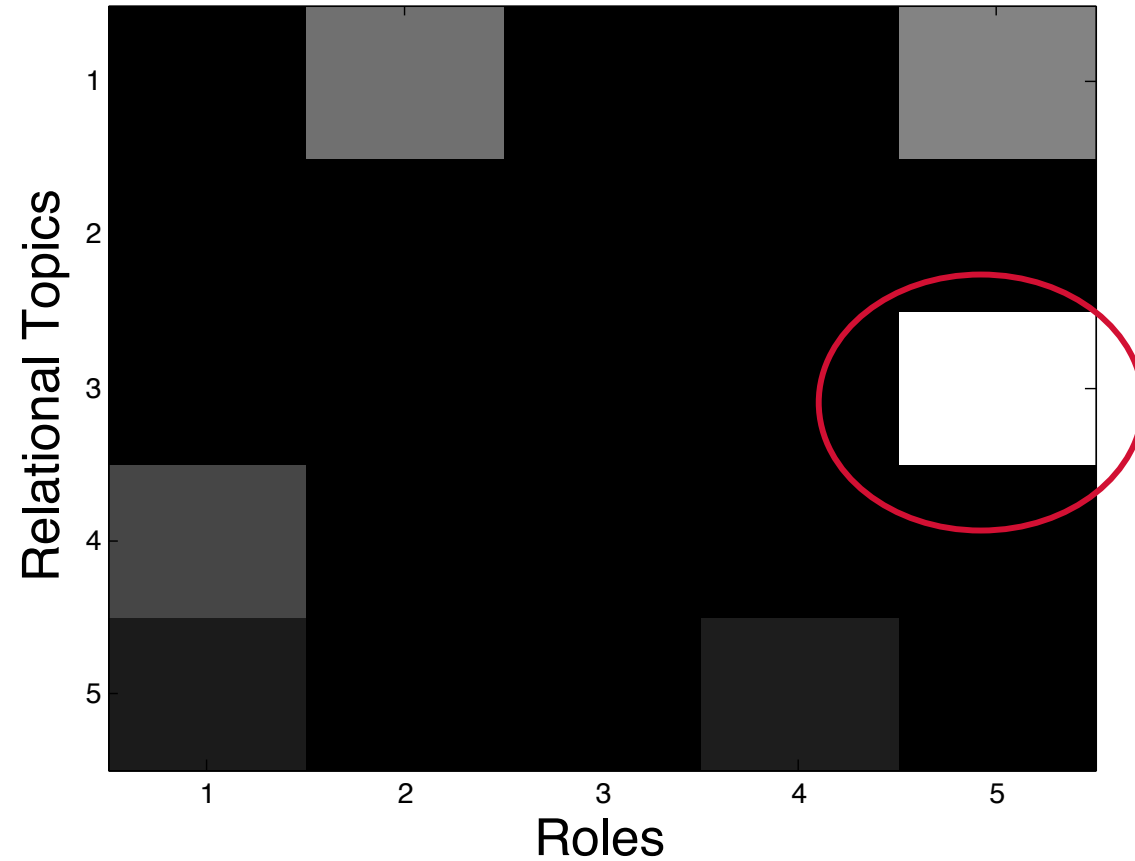
More insights into Group 1

Group 1

Name	Party	Exp
Millender-McDonald	D	11
Obey, David	D	38
Tsongas, Niki	D	0
Speier, Jackie	D	0
Faleomavaega, Eni	D	18
Meehan, Martin	D	14
Edwards, Donna	D	0
Visclosky, Peter	D	22
Hoyer, Steny	D	26
Foster, Bill	D	0

Relational Topic 3
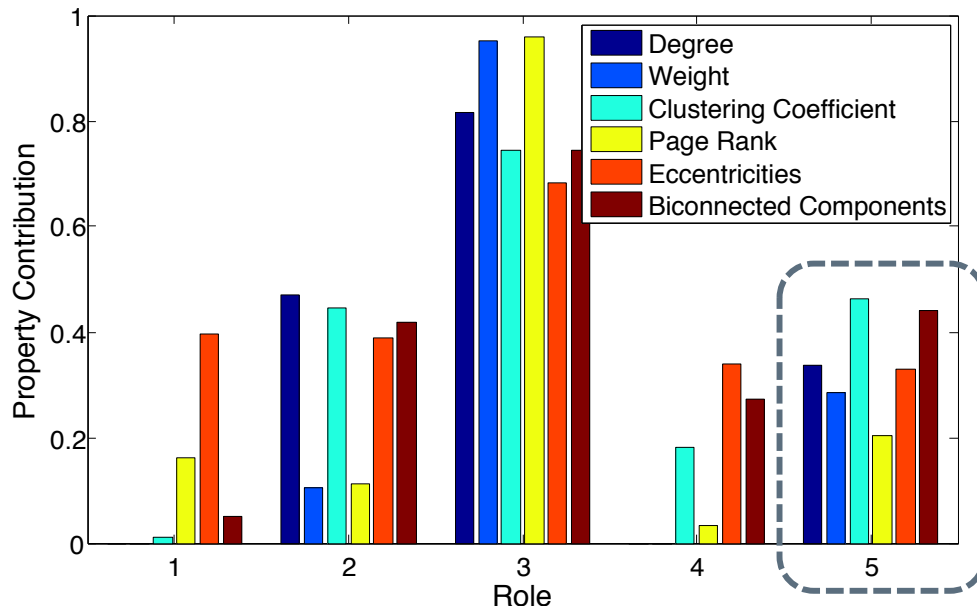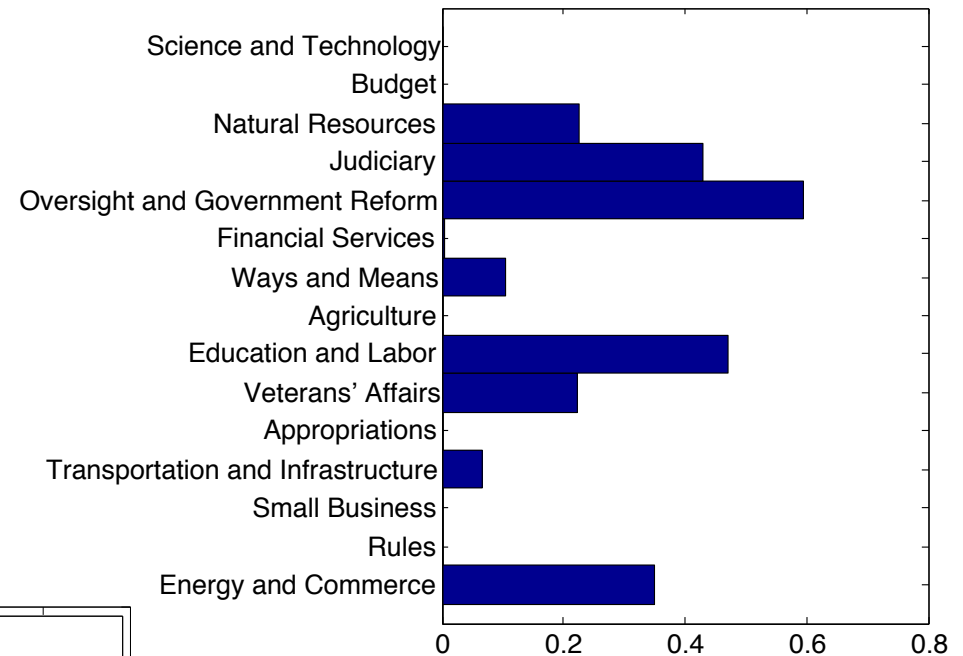
Group 1

- Democrats; mostly not mid-career
- Active in oversight & gov't reform
- On the periphery, but lots of triangles

# Multi-relational Role Discovery (MRD)

[Gilpin et al., ArXiv 2016]
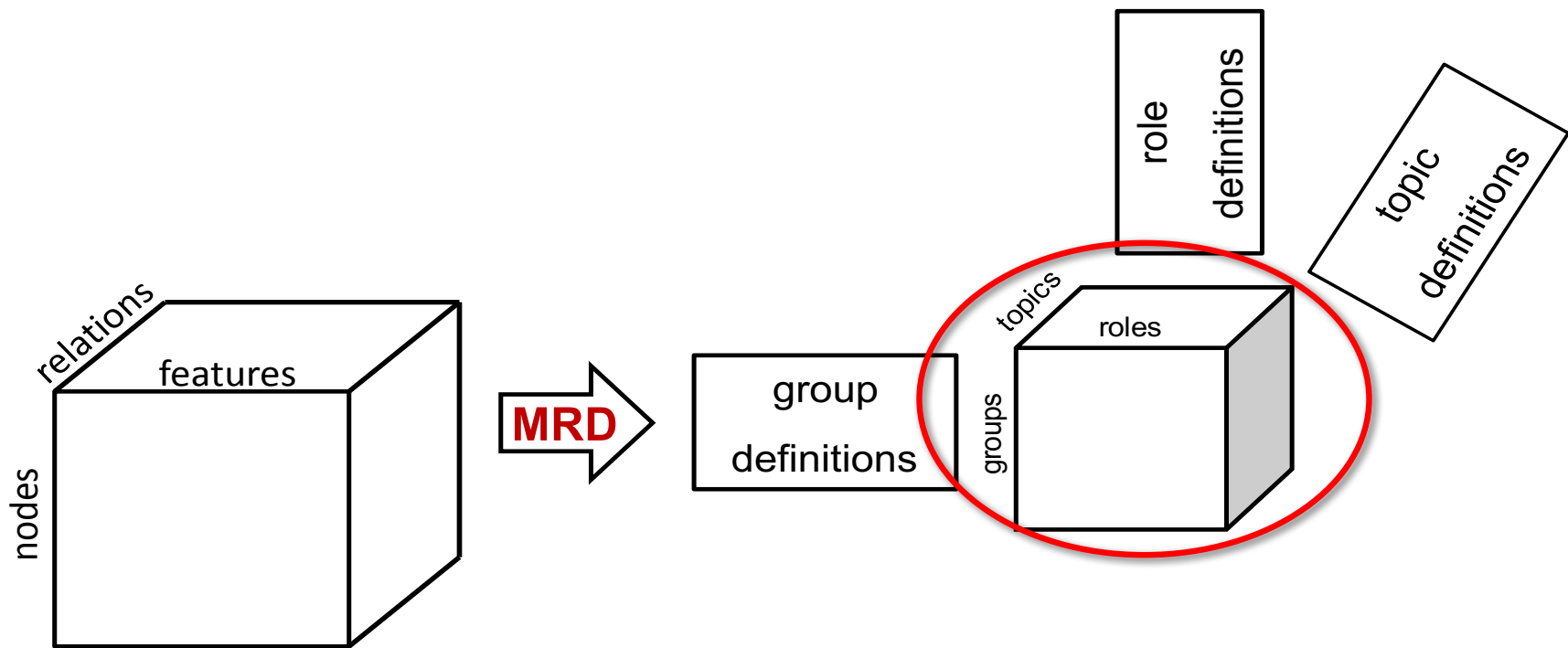
Interaction Graph of a Tucker Core
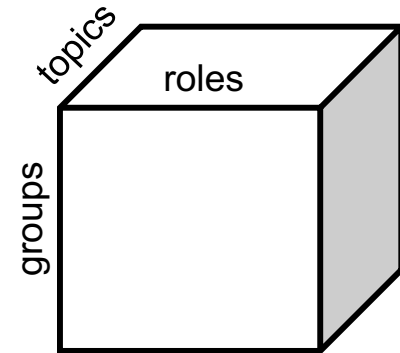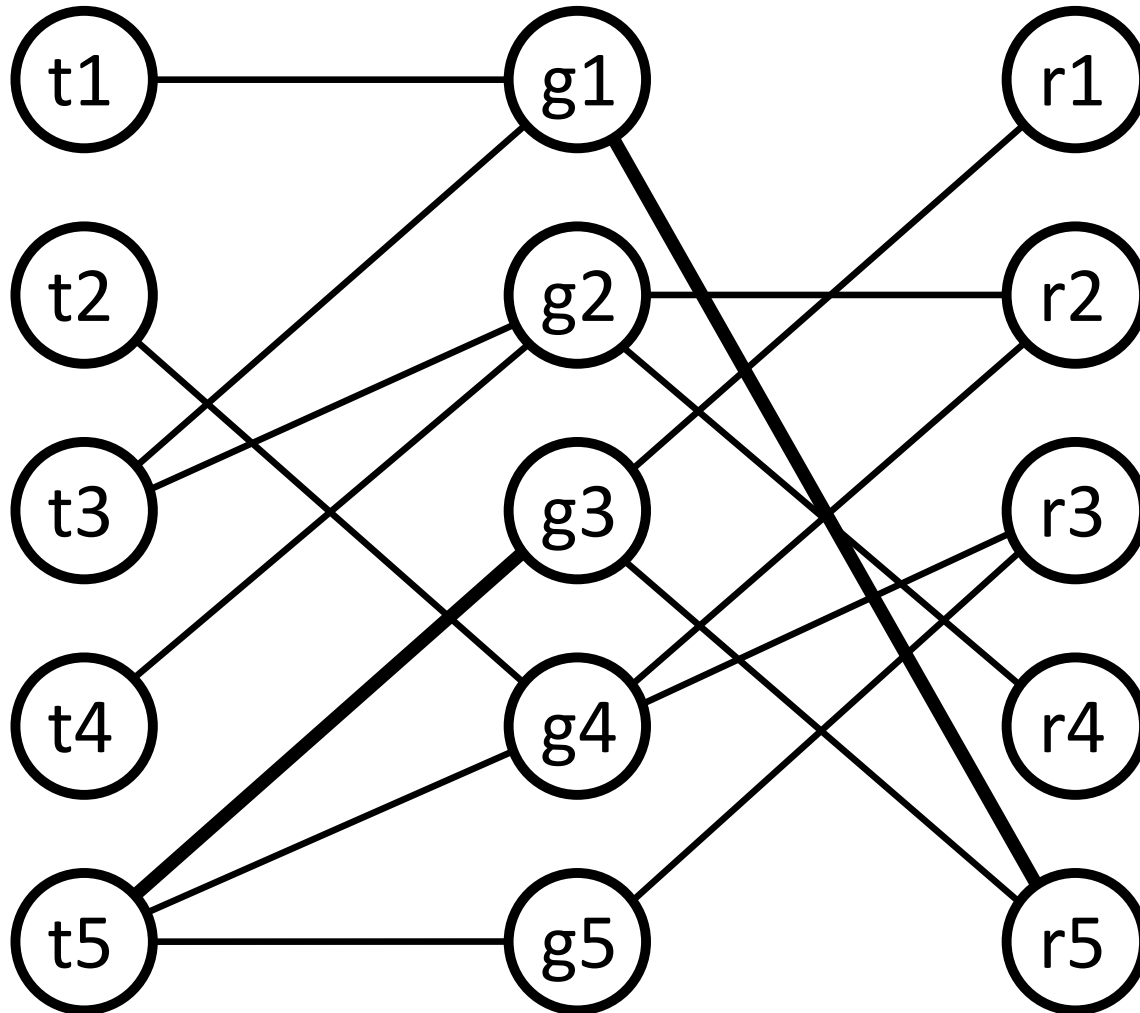
From the
110th Congress
Co-sponsorships

Measure Properties on the Interaction Graph

| Property	Description	Computation
Simplicity	To what extent are nodes connected to (role) similar types of nodes?	Average Node Degree
Sharing	How much can a group be separated into independent parts?	Mincut cost
Variability	How does the simplicity of nodes vary across the interaction graph?	Variance of node degree
Stability	How stable are the interactions between roles, groups, and topics?	Spectral gap

Macroscopic Properties of the Interaction Graphs from Tucker Cores
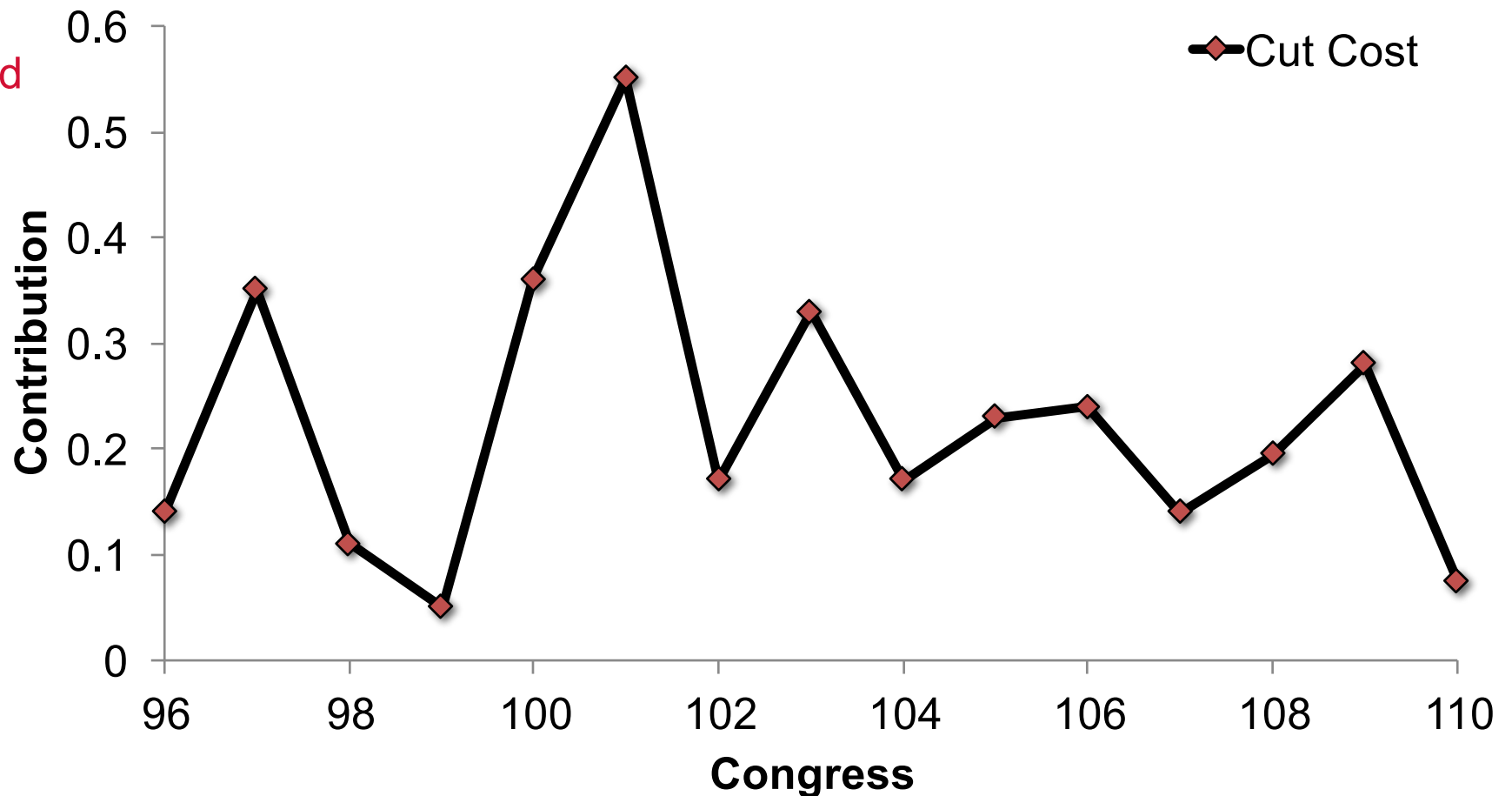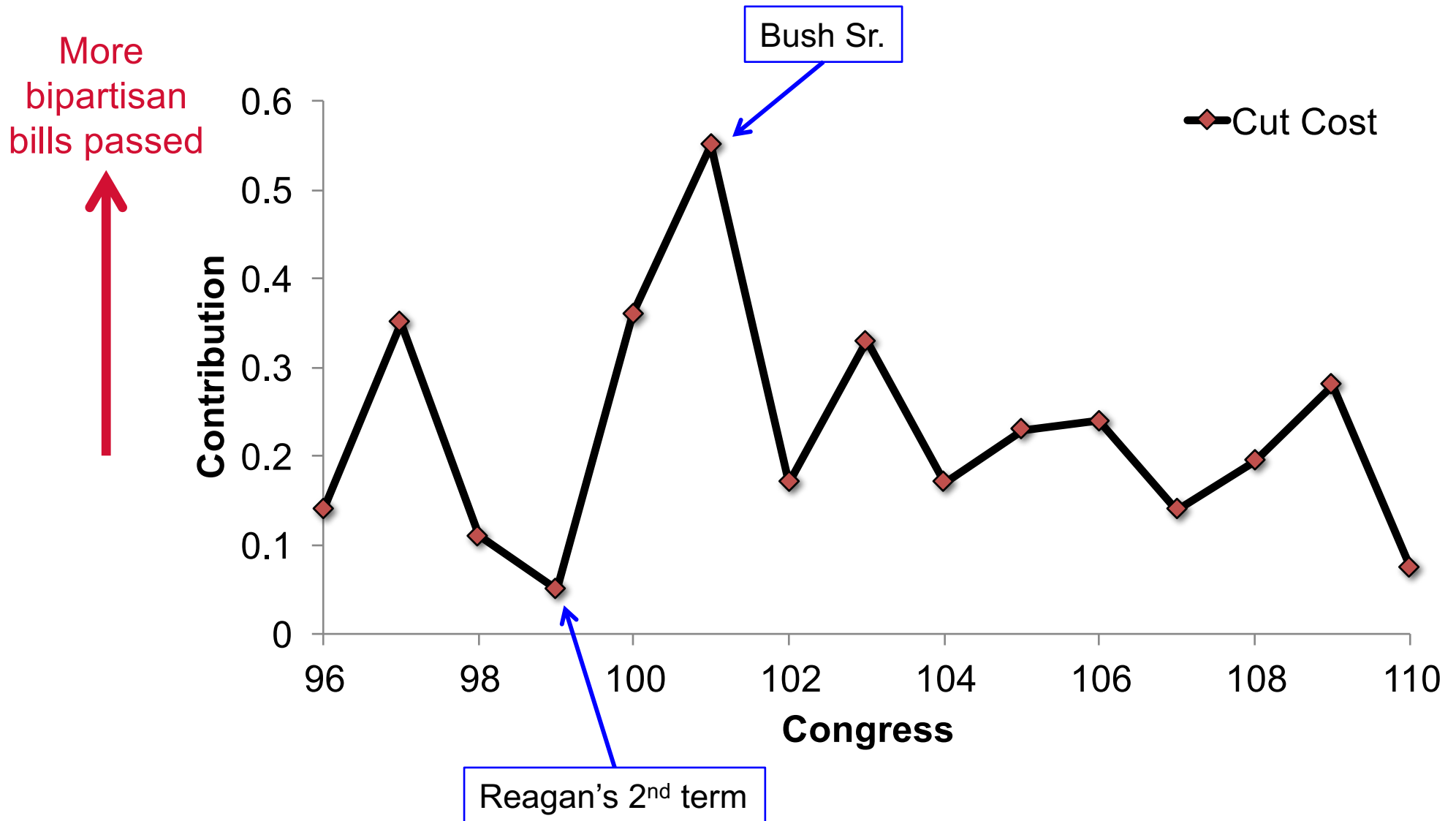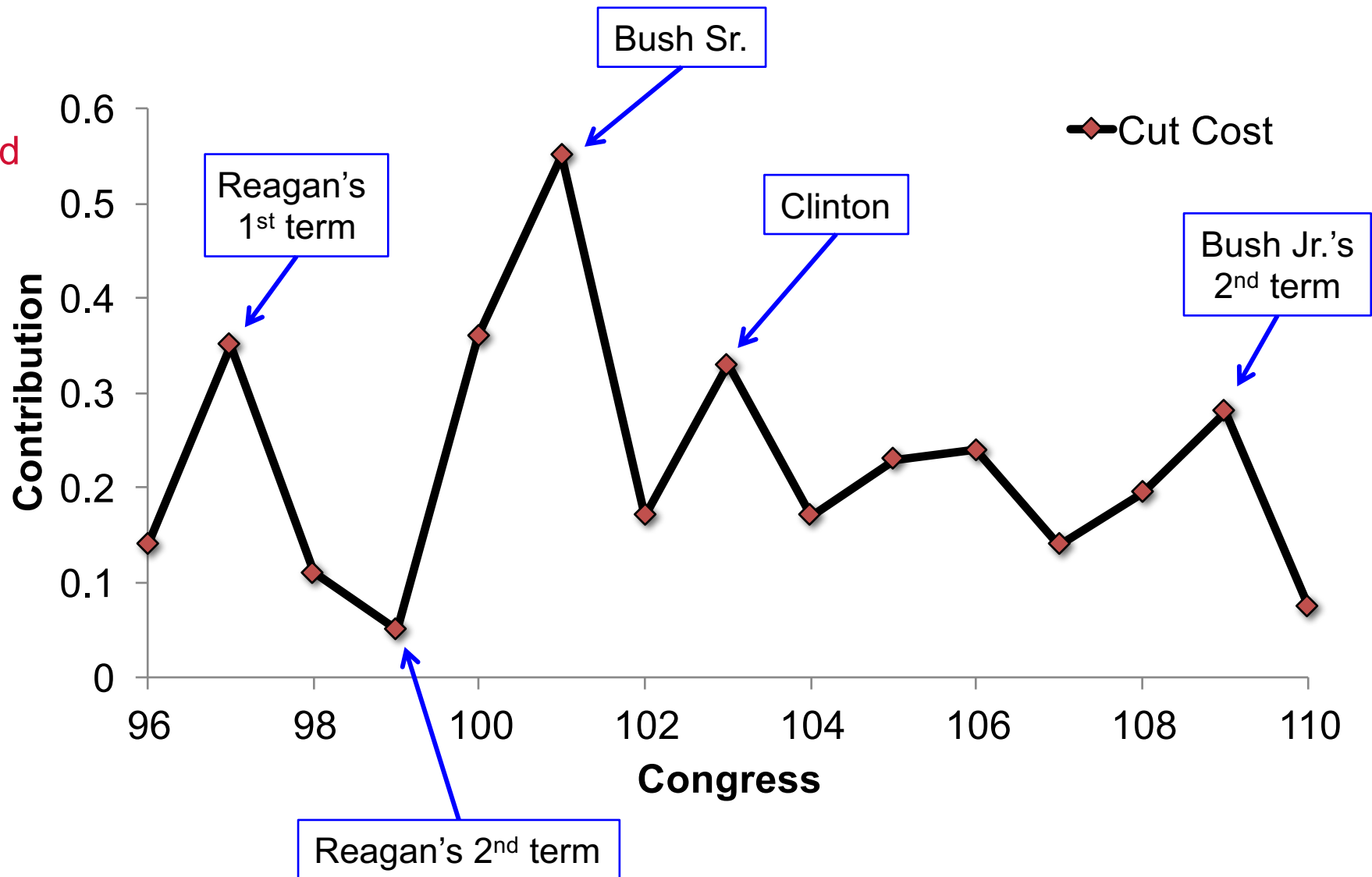
Macroscopic Properties of the Interaction Graphs from Tucker Cores
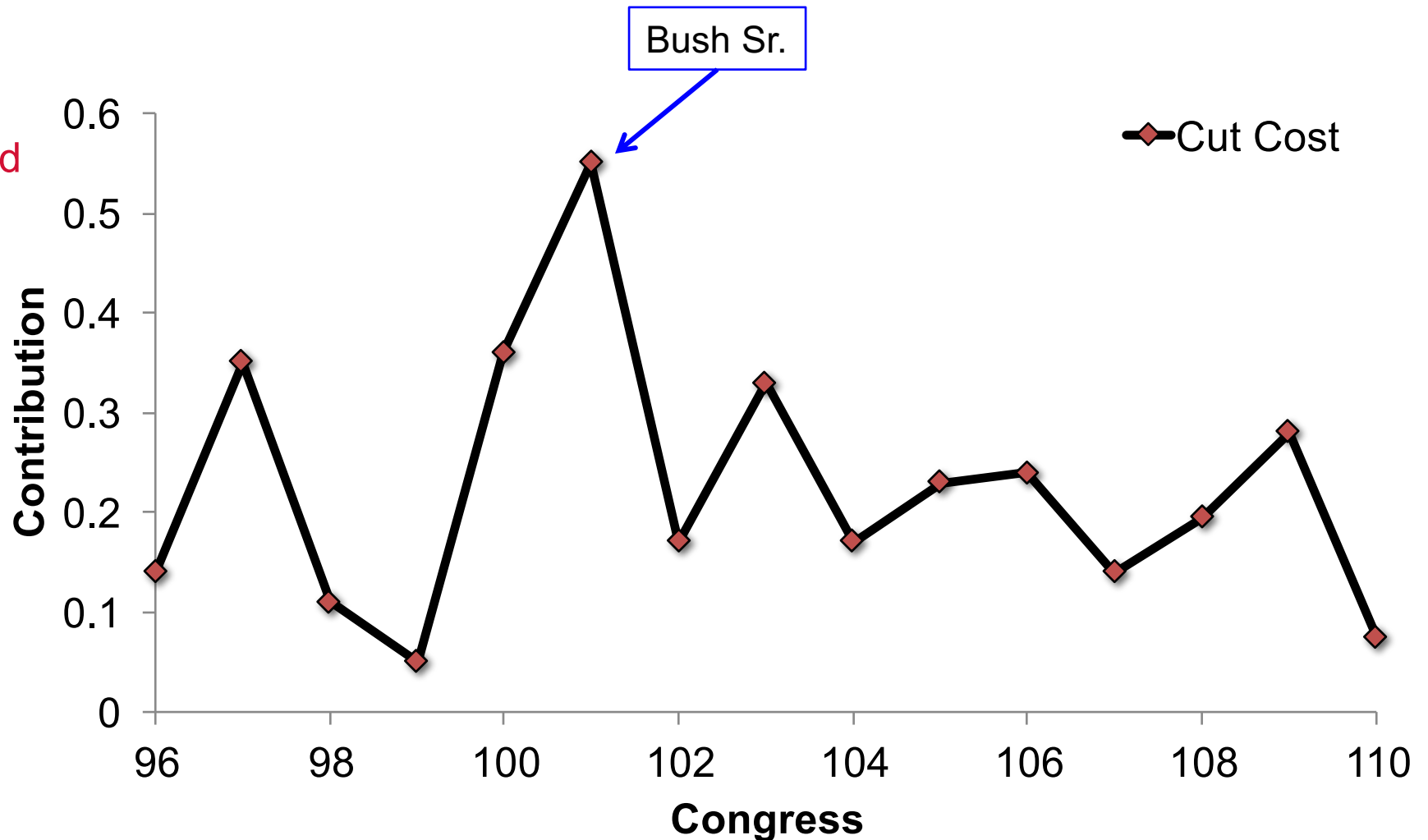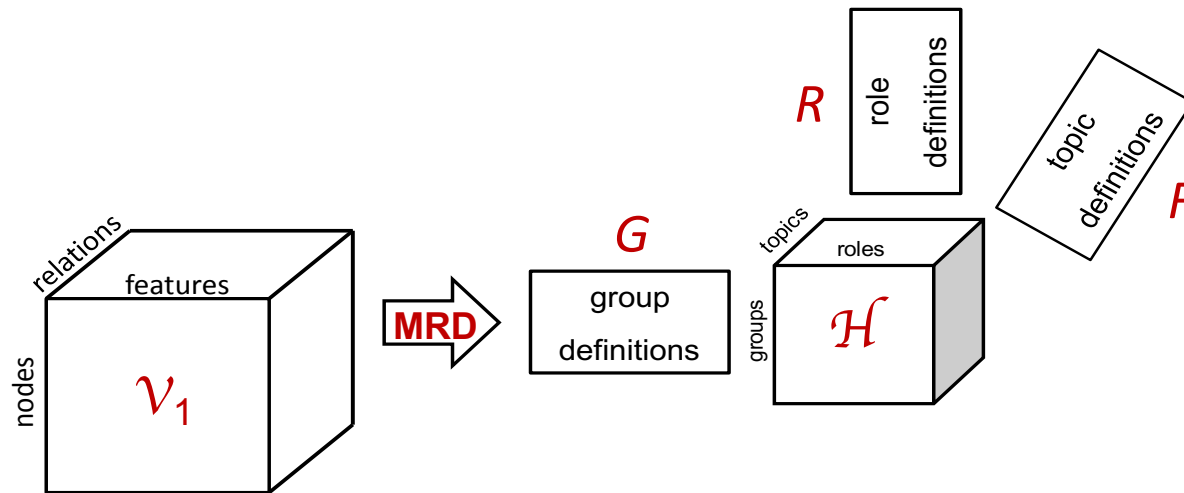
Macroscopic Properties of the Interaction Graphs from Tucker Cores

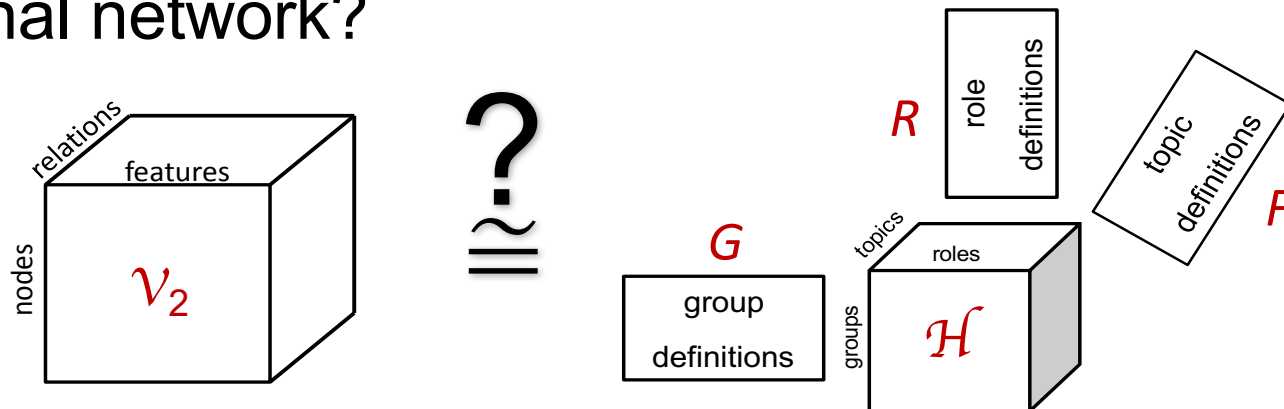# Macroscopic Properties of the Interaction Graphs from Tucker Cores

Role Transfer in MRD

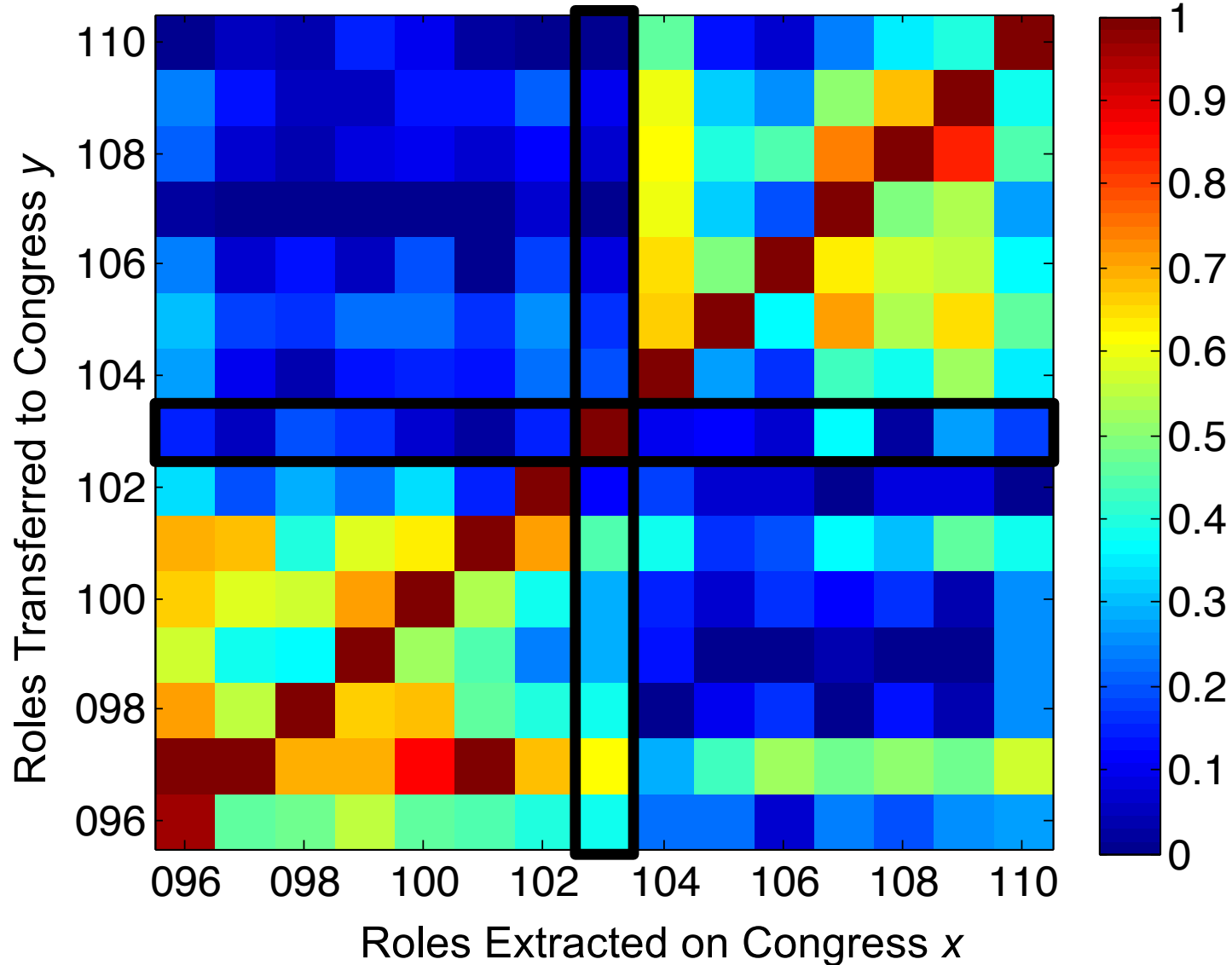- Extract roles on one multi-relational network

- How well do the extracted roles transfer to another multi-relational network?

Role Transfer on Multi-relational Networks

Heatmap of fit quality = 1 – normalized reconstruction error

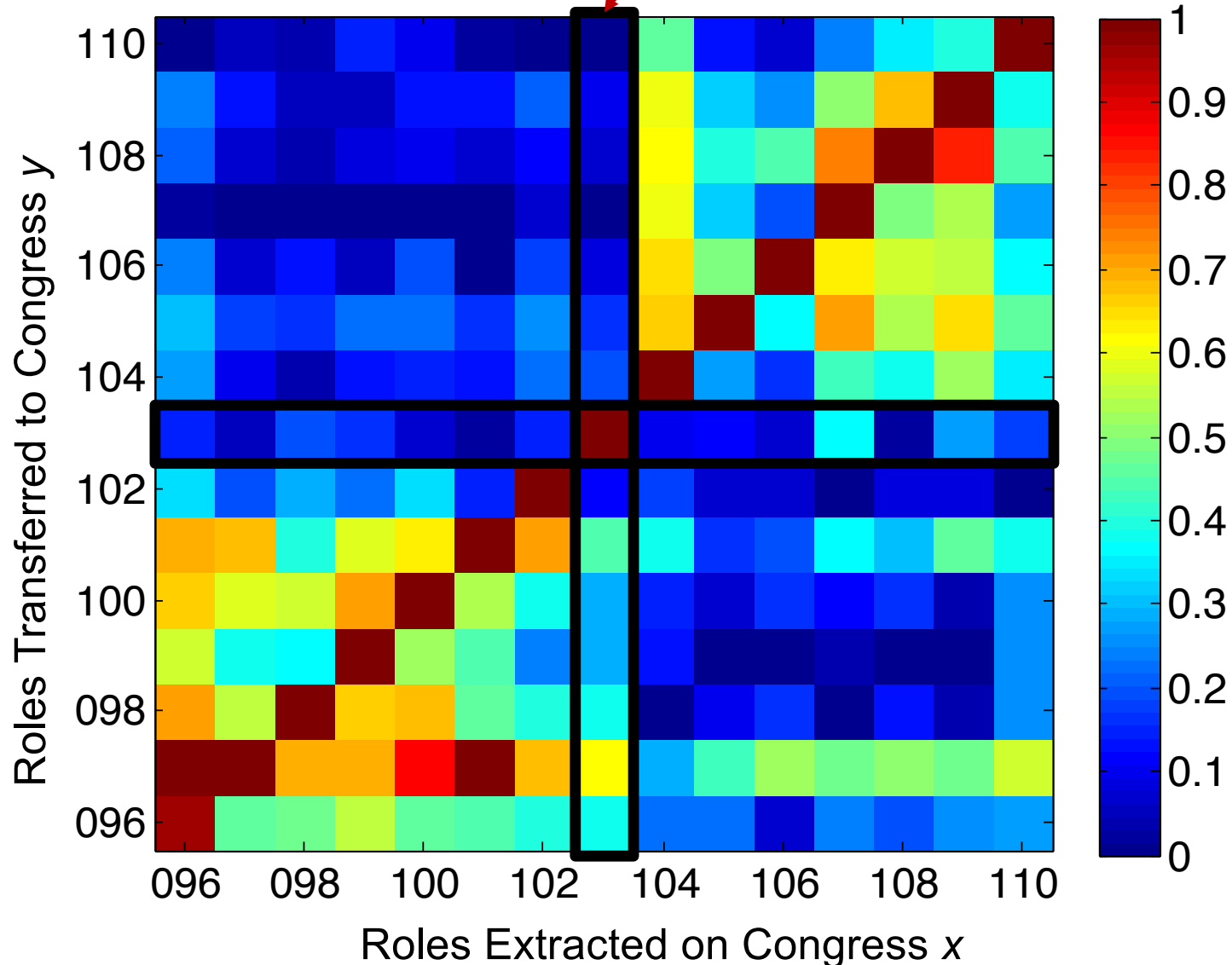# Role Transfer on Multi-relational Networks

Heatmap of fit quality = 1 – normalized reconstruction error

Role Transfer

Hastert Rule: the Speaker will not allow a floor vote on a bill unless a majority of the majority party supports the bill.

Heatmap of fit quality = 1 – normalized reconstruction error

Why are Roles Effective in Many Applications?

- Encode complex behavior

- Map nodes into a useful lower dimensional space

- Generalize across networks

Lots more to do …

- An in-depth study on properties of these latent role spaces

- Information spread through roles

  - How roles affect influence & susceptibility?

Lots more to do …

- An in-depth study on properties of these latent role spaces

- Information spread through roles

  - How roles affect influence & susceptibility?

- Combining physics of networks (PoN) with the mining of graphs (MoG)

  - What are the functional roles in an ensemble of networks?

  - How do we incorporate functional roles from instances of networks into PoN models?

Thank You!

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

- Tutorials at http://eliassi.org

- Open-source code at https://snap.stanford.edu/snap-2.3/

- Joint work with

  - LLNL (Keith Henderson & Brian Gallagher)

  - CMU (Christos Faloutsos, Leman Akoglu et al.)

  - Google (Sugato Basu)

  - UC Davis (Ian Davidson et al.)