



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

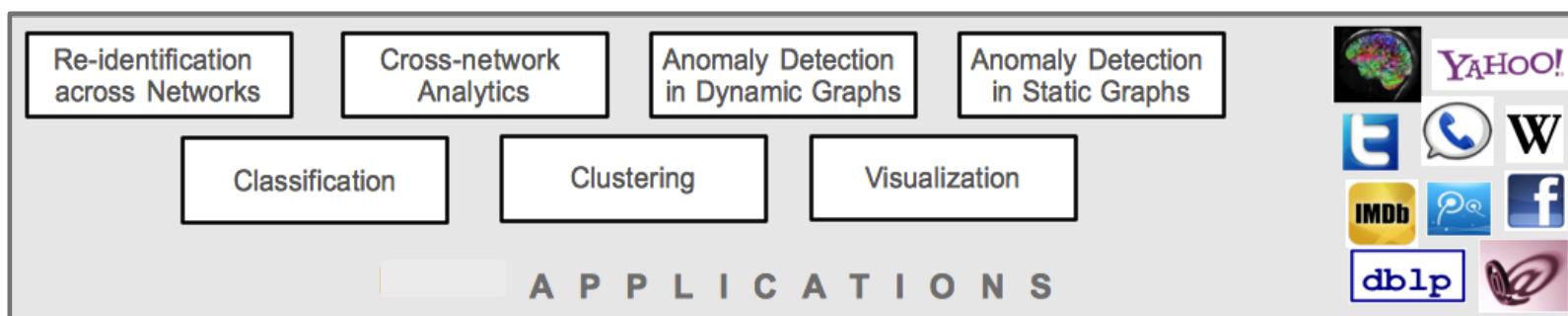
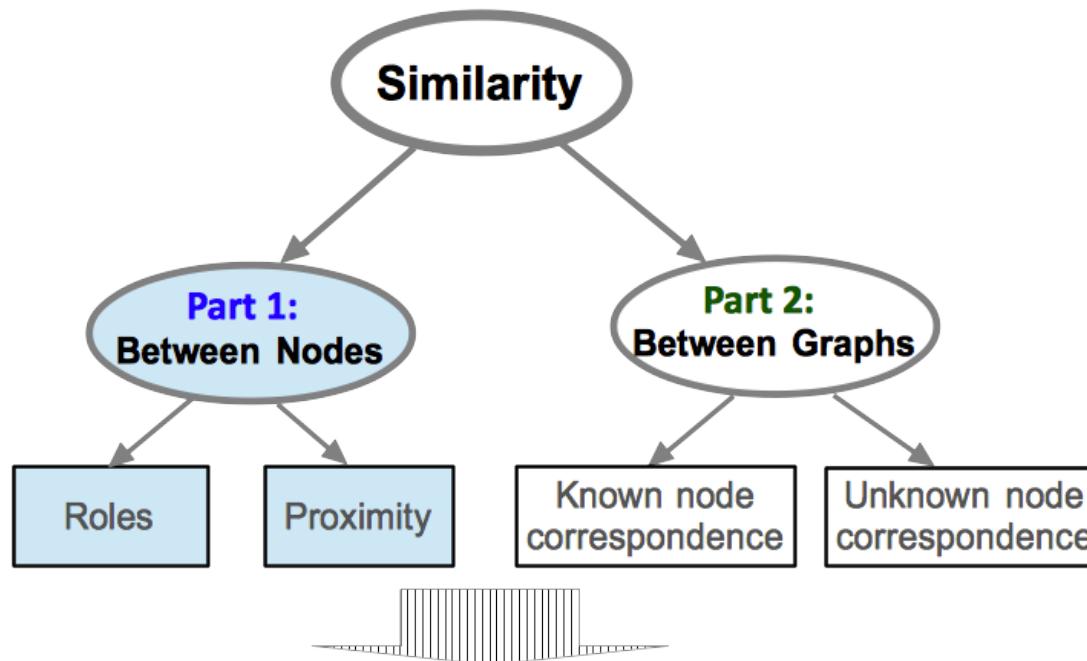
Copyright for the tutorial materials is held by the authors. The authors grant IEEE ICDM permission to distribute the materials through its website.

Who we are

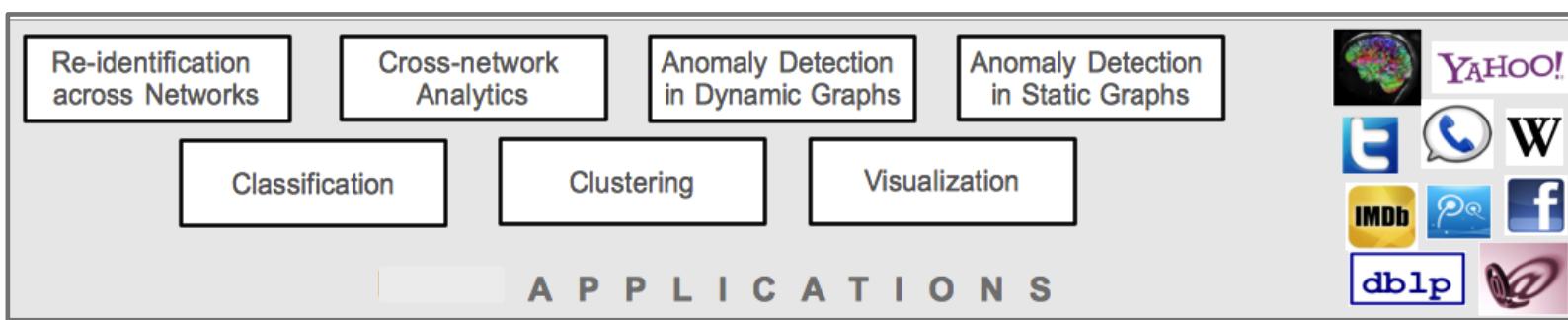
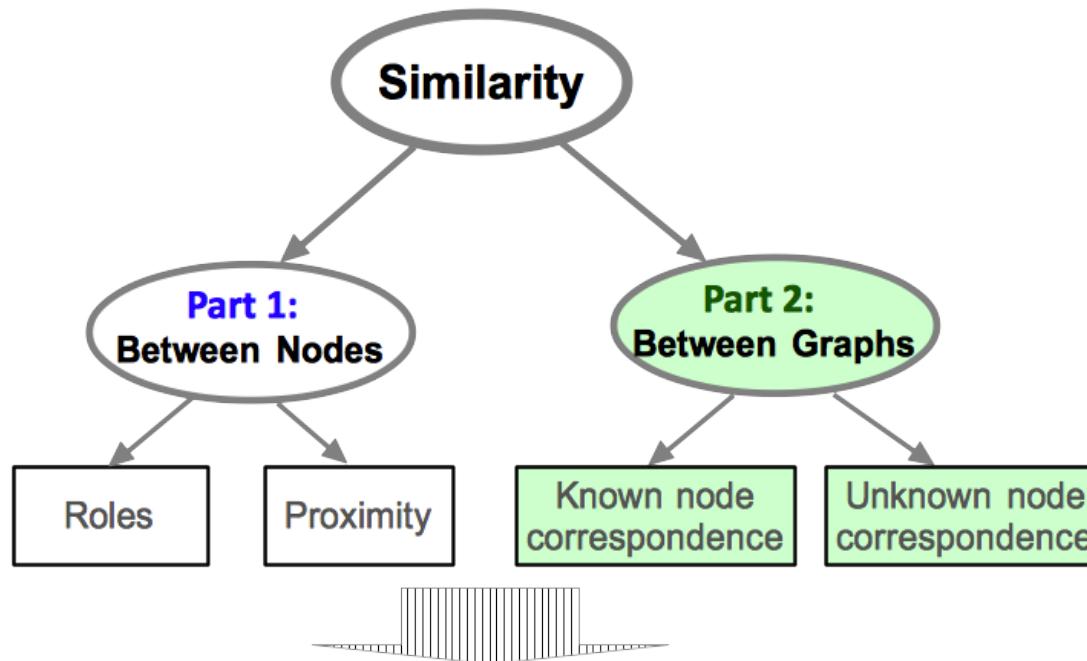
- Danai Koutra, CMU
 - Node and graph similarity, summarization, pattern mining
 - <http://www.cs.cmu.edu/~dkoutra/>
- Tina Eliassi-Rad, Rutgers
 - Data mining, machine learning, big complex networks analysis
 - <http://eliassi.org/>
- Christos Faloutsos, CMU
 - Graph and stream mining, ...
 - <http://www.cs.cmu.edu/~christos>



What we will cover

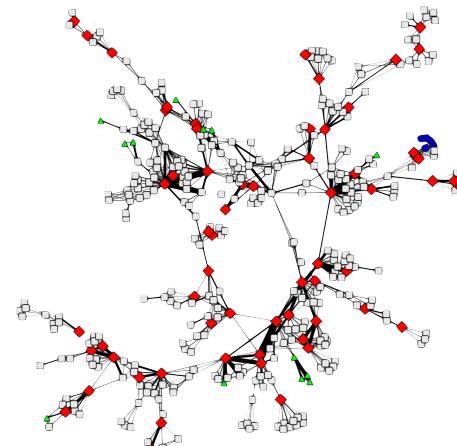
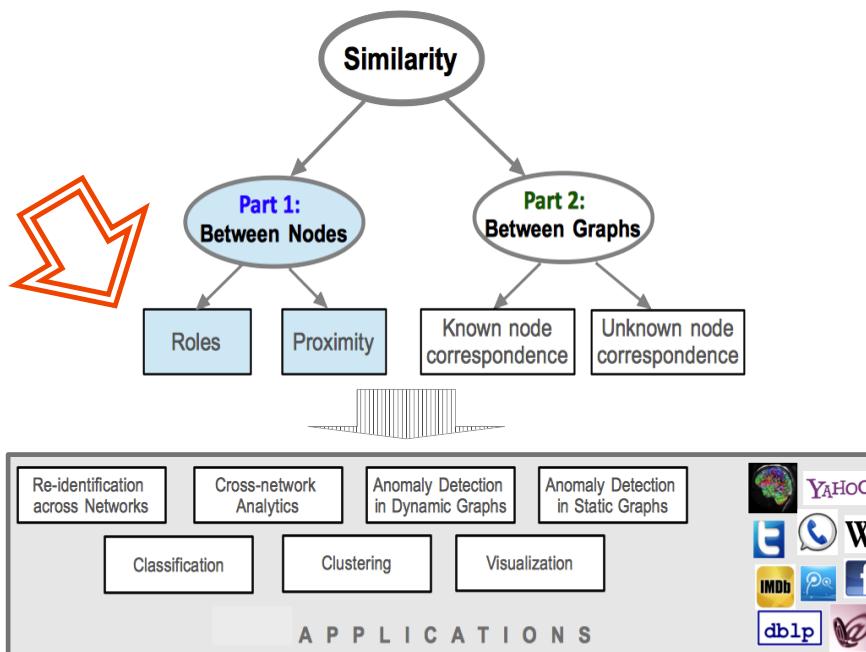


What we will cover



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)

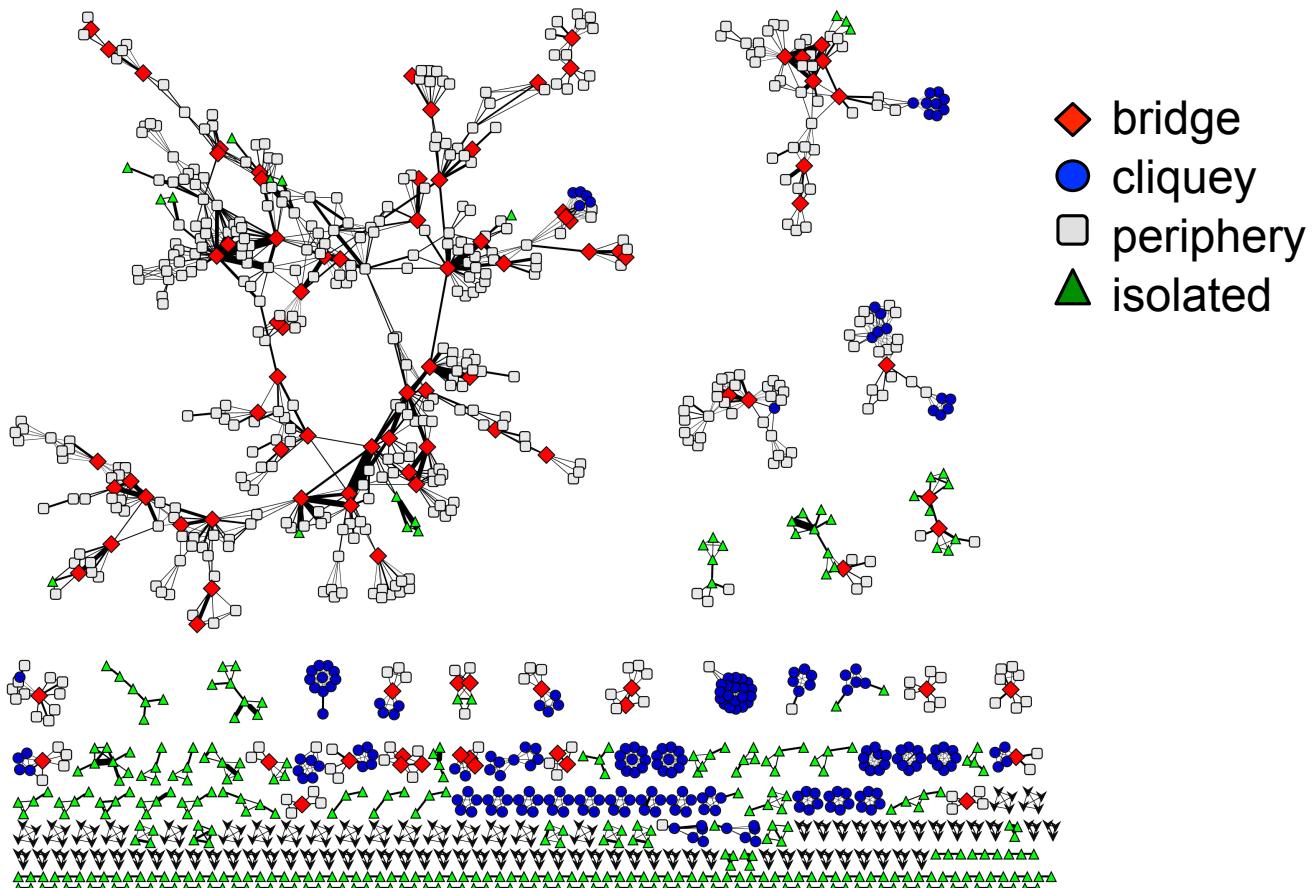


What are roles?

- “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

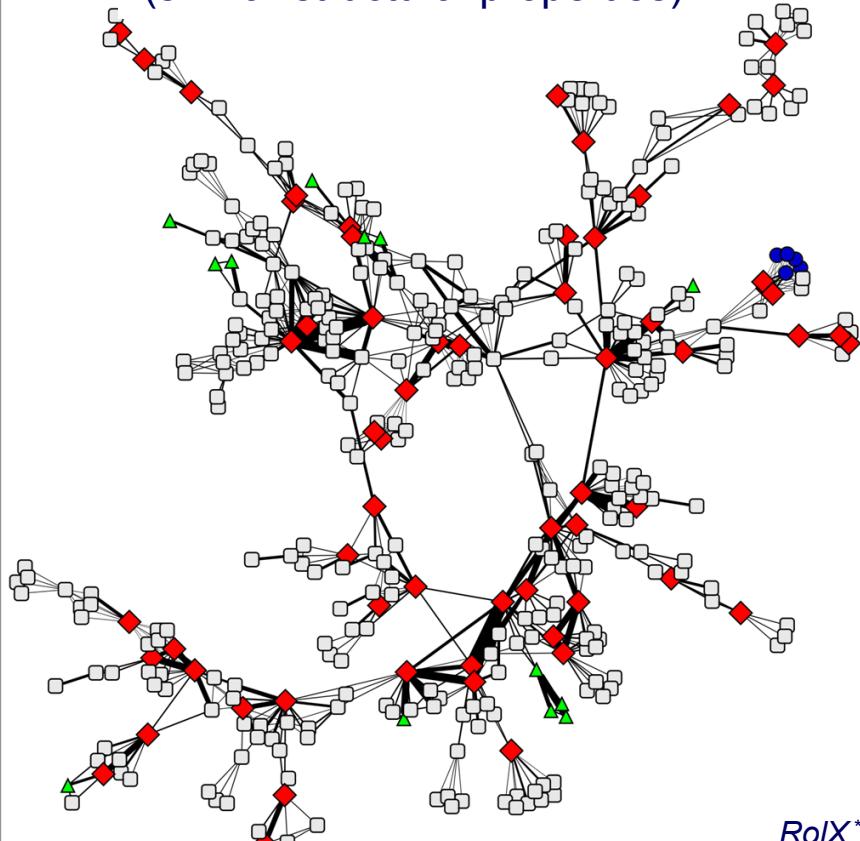
- Node Roles
 - What are roles
 - Roles and communities
 - Roles and equivalences (from sociology)
 - Roles (from data mining)
 - Summary
- Node Proximity



Roles and Communities are Complementary

Roles

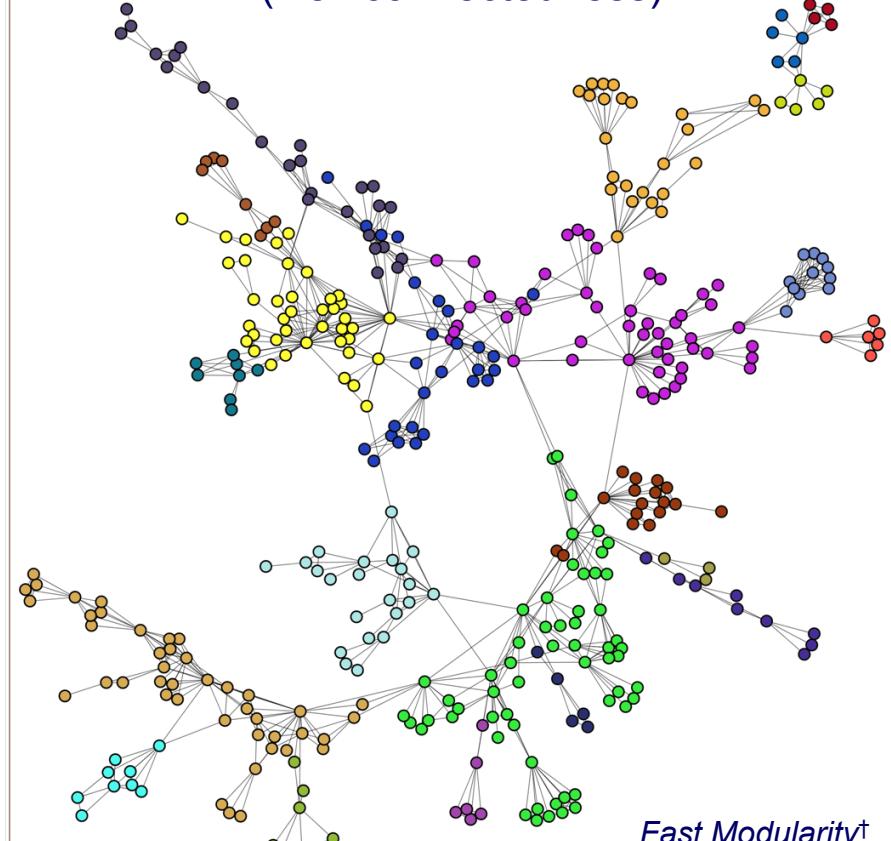
(similar structural properties)



*RoX**

Communities

(well-connectedness)



Fast Modularity†

* 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

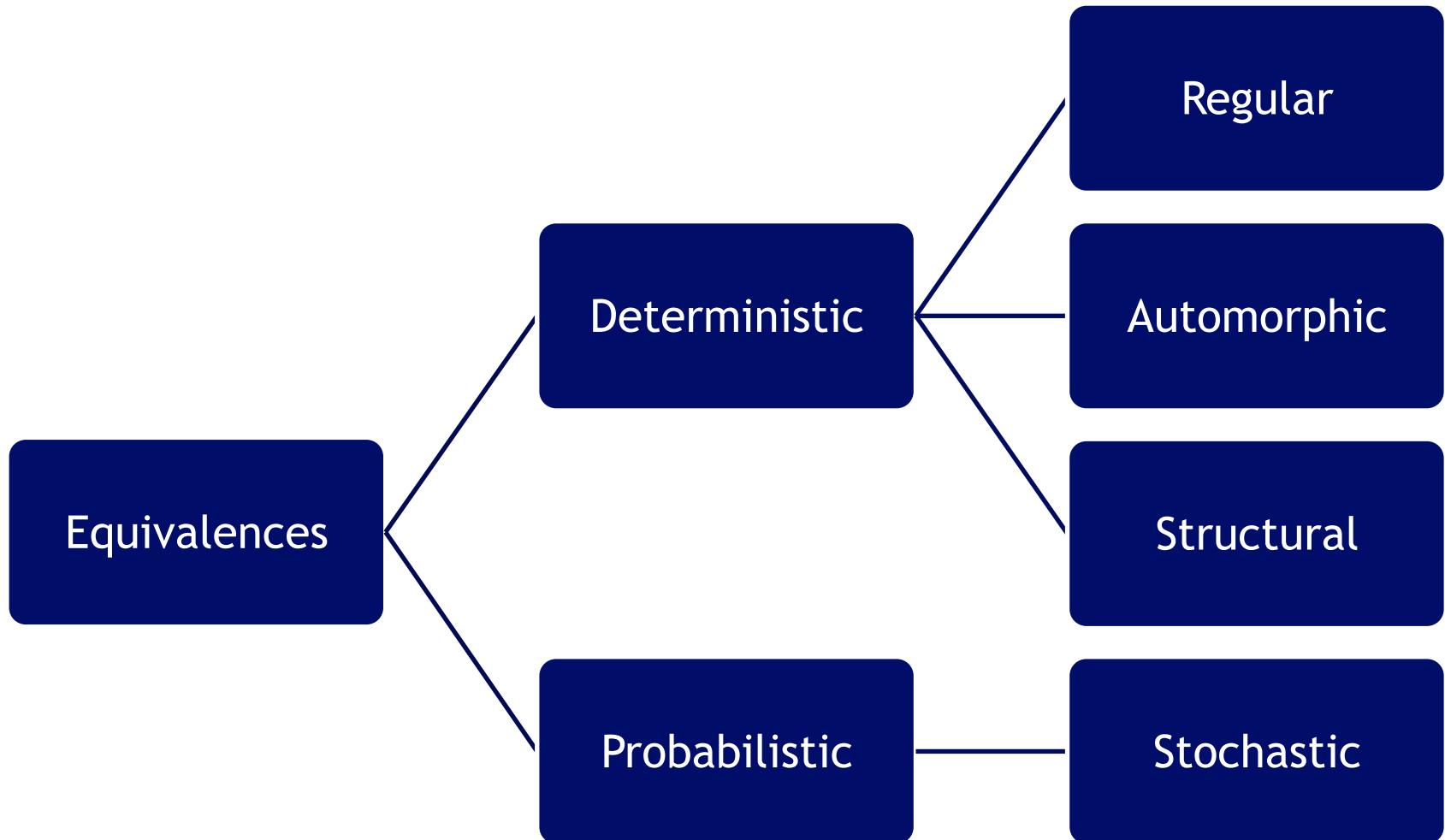


Roles == Positions in Sociology

- 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

Regular

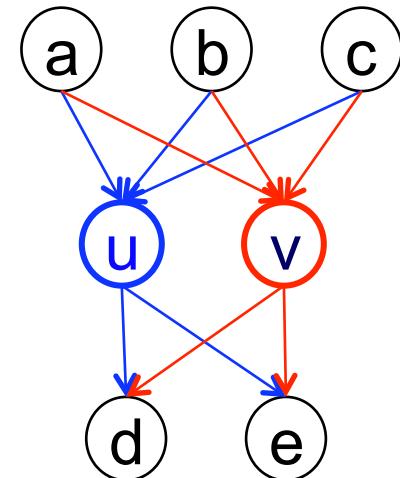
Automorphic

Structural



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

Regular

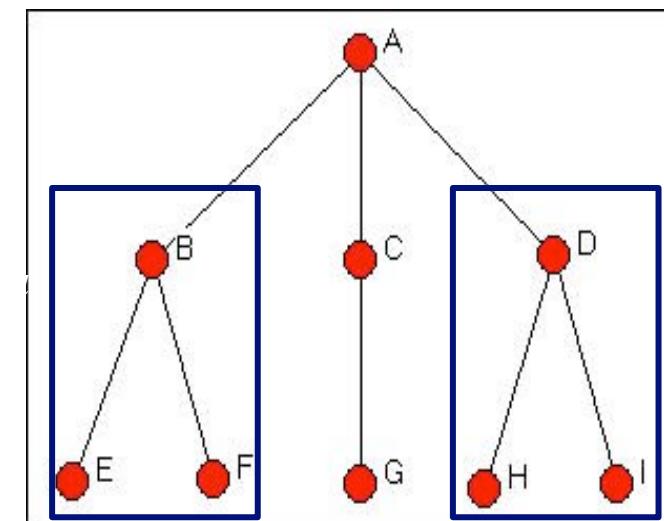
Automorphic

Structural



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

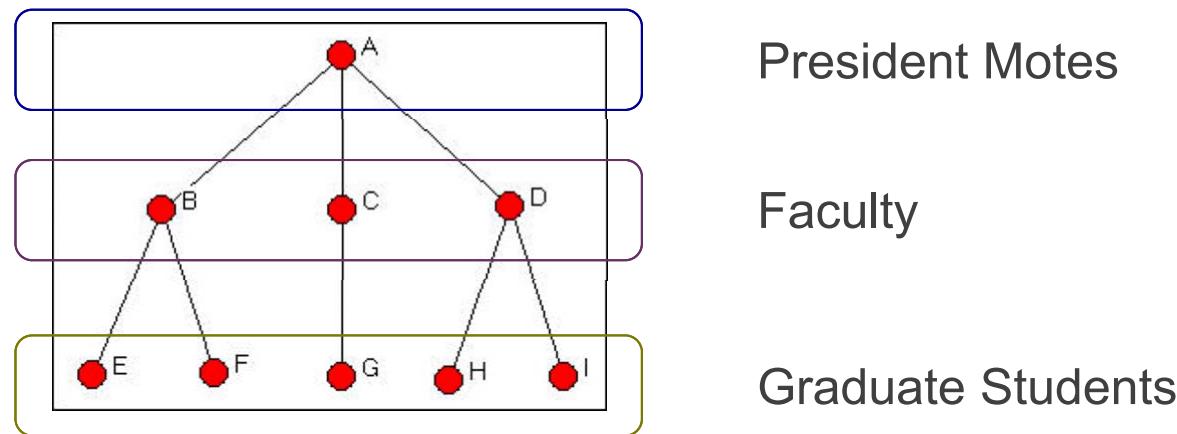
Automorphic

Structural



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/>)

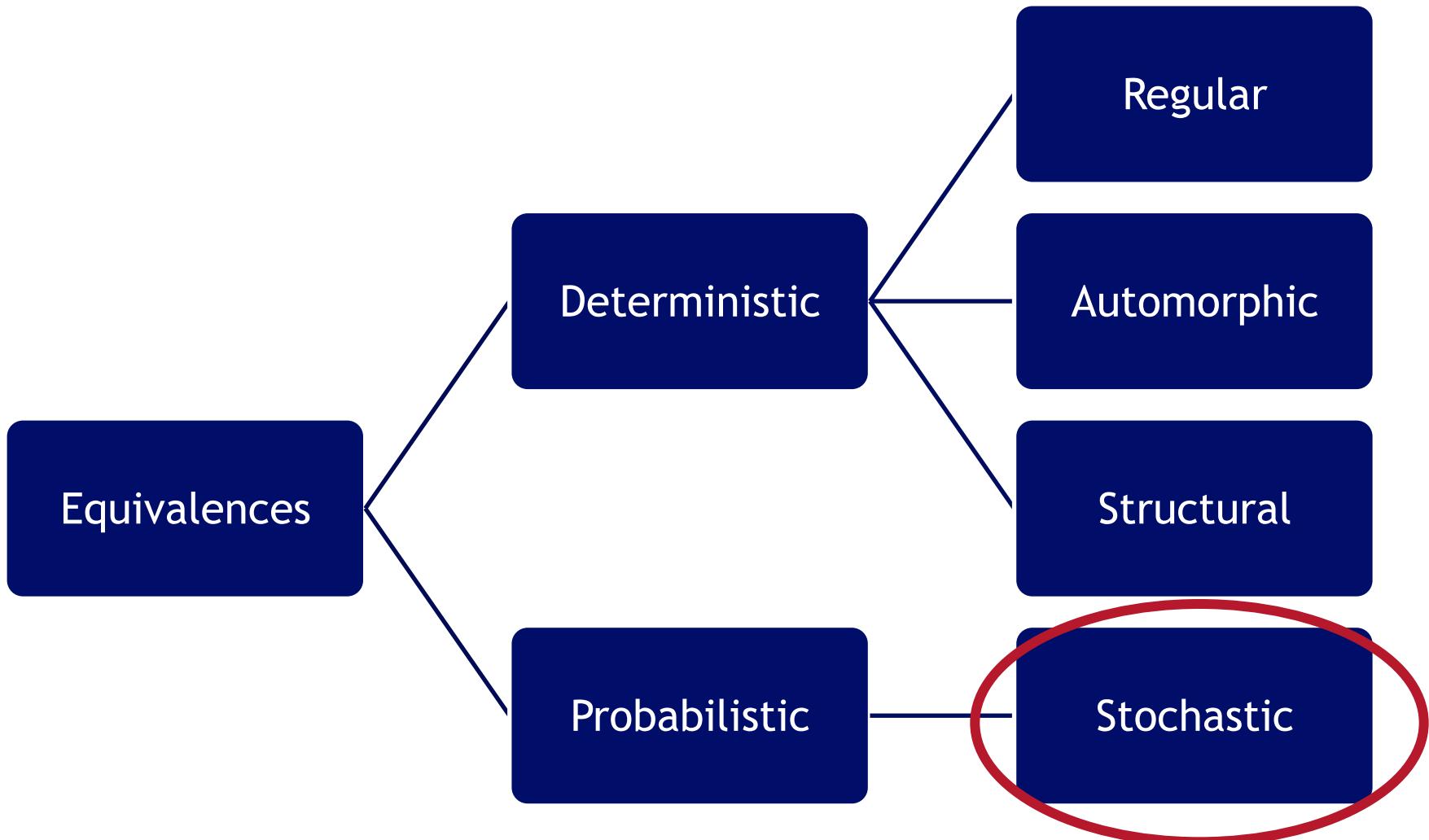


Examples of Algorithms for Deterministic Equivalences

- 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

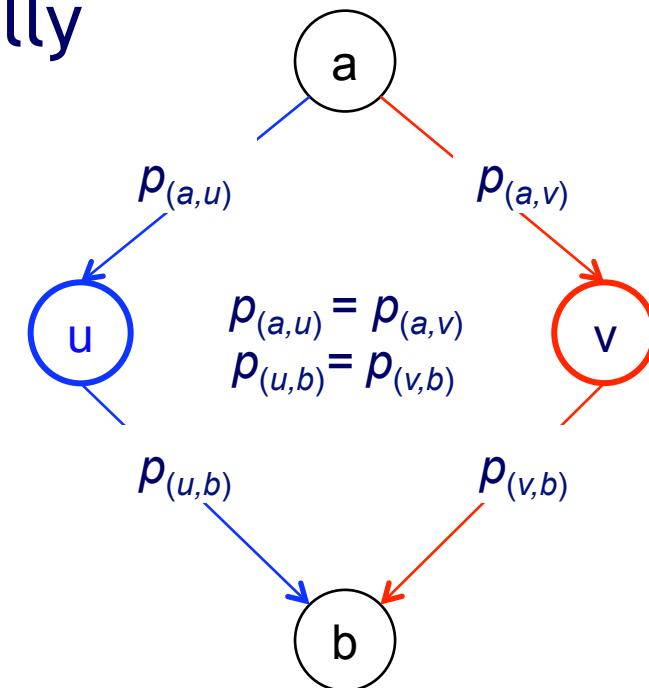


Equivalences



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 <i>et al.</i> , KDD 2011]
RoLX	Role Extraction from Structural Features	[Henderson <i>et al.</i> , KDD 2012]
GLRD	Guided Learning for Role Discovery	[Gilpin <i>et al.</i> , KDD 2013]
DBMM	Dynamic Behavioral Mixed-membership Model	[Rossi <i>et al.</i> , WSDM 2013]
MRD	Multi-relational Role Discovery	[Gilpin <i>et al.</i> , in preparation]
LearnMelt	Learning to Minimize Information Dissemination	[Le <i>et al.</i> , in preparation]
...		

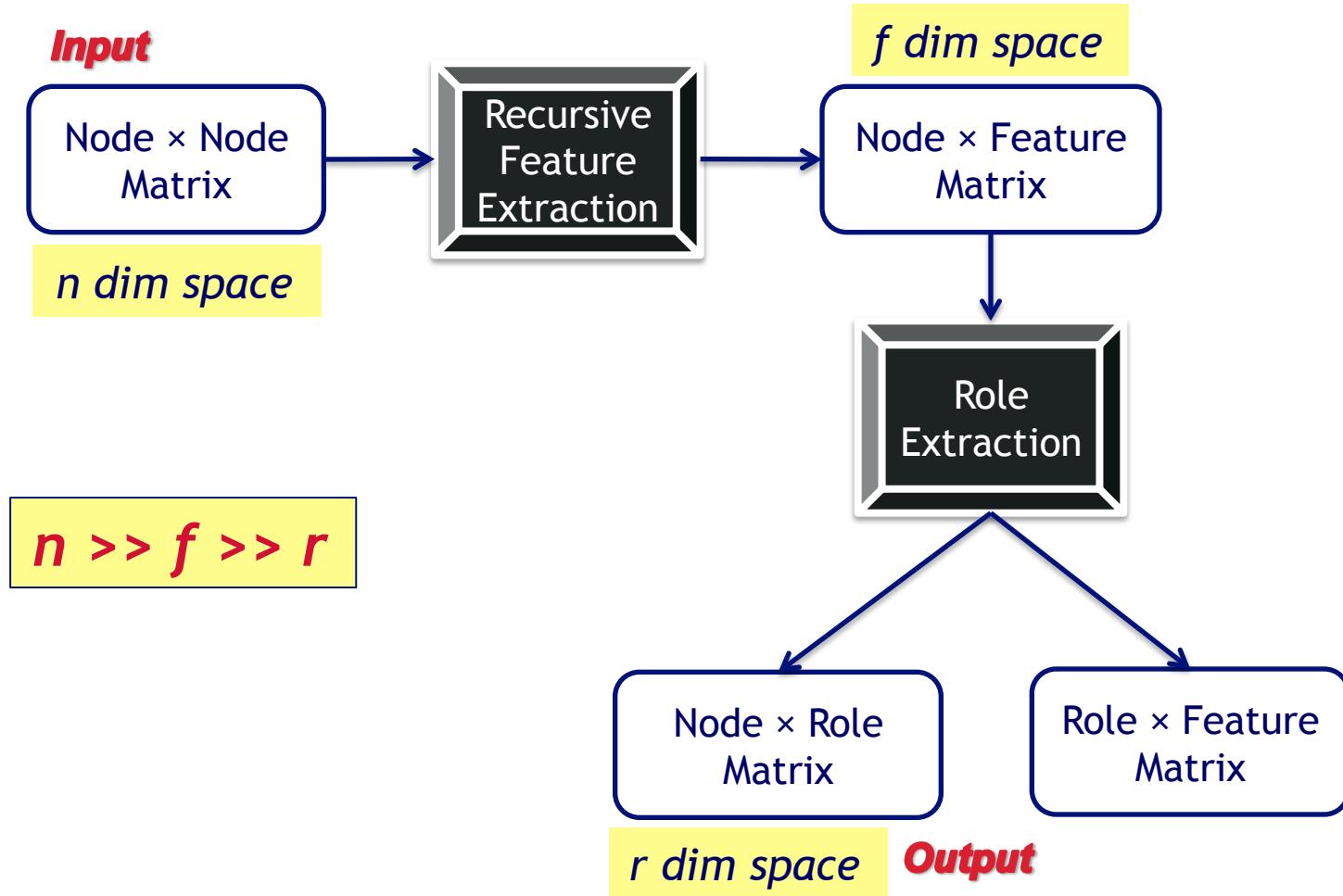


RolX: Role eXtraction

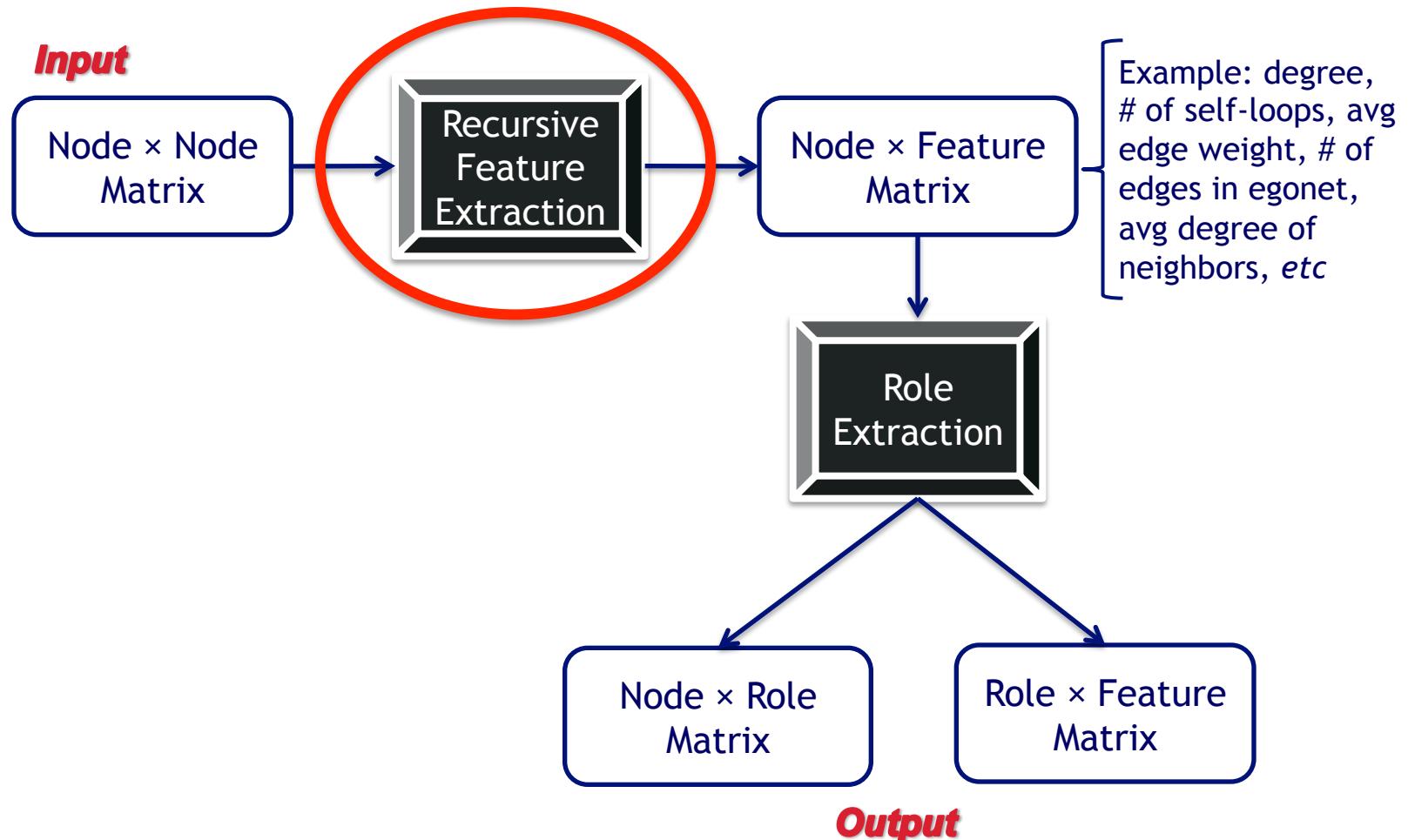
- Introduced by Henderson et al. KDD 2012
- Automatically extracts the underlying roles in a network
 - No prior knowledge required
- Determines the number of roles automatically
- Assigns a mixed-membership of roles to each node
- Scales linearly on the number of edges

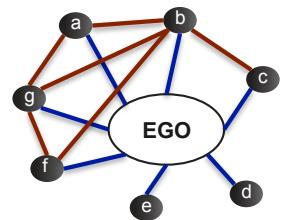


RolX: Flowchart



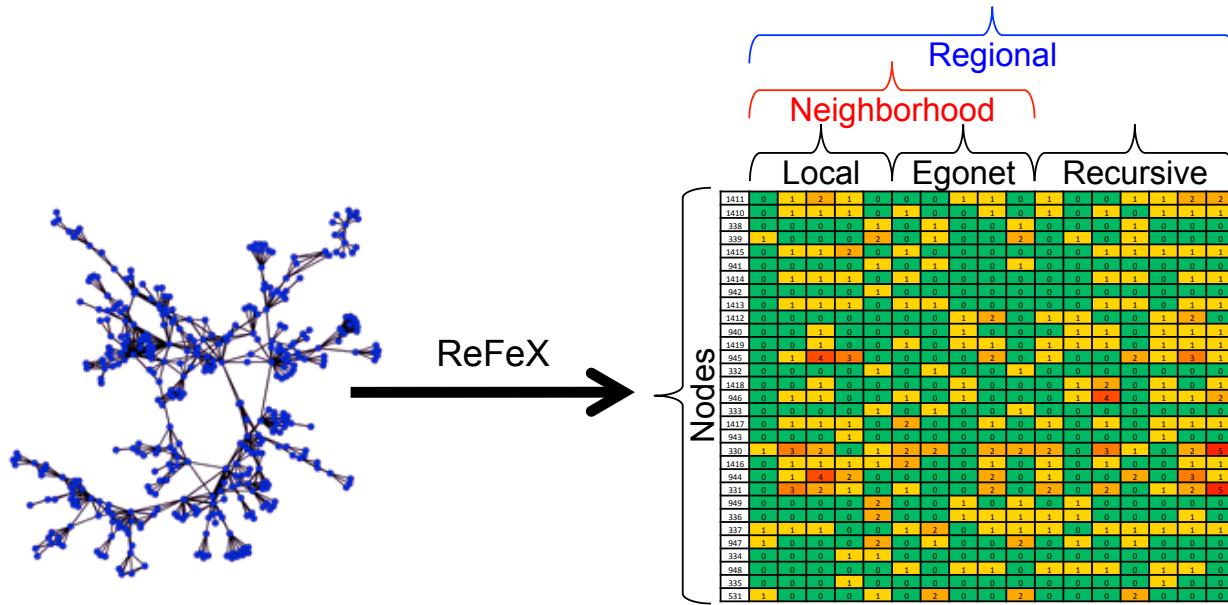
RolX: Flowchart





Recursive Feature Extraction

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

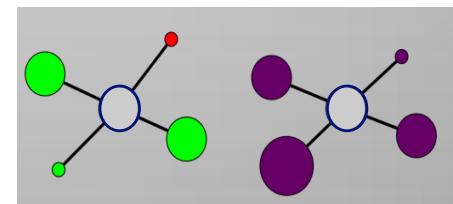
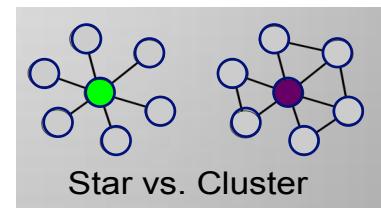
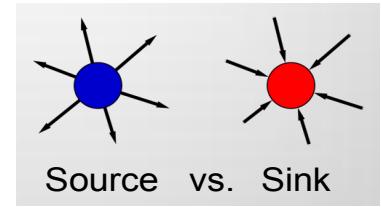
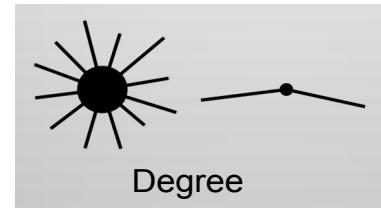


- Neighborhood features: What is your connectivity pattern?
- Recursive Features: To what kinds of nodes are you connected?

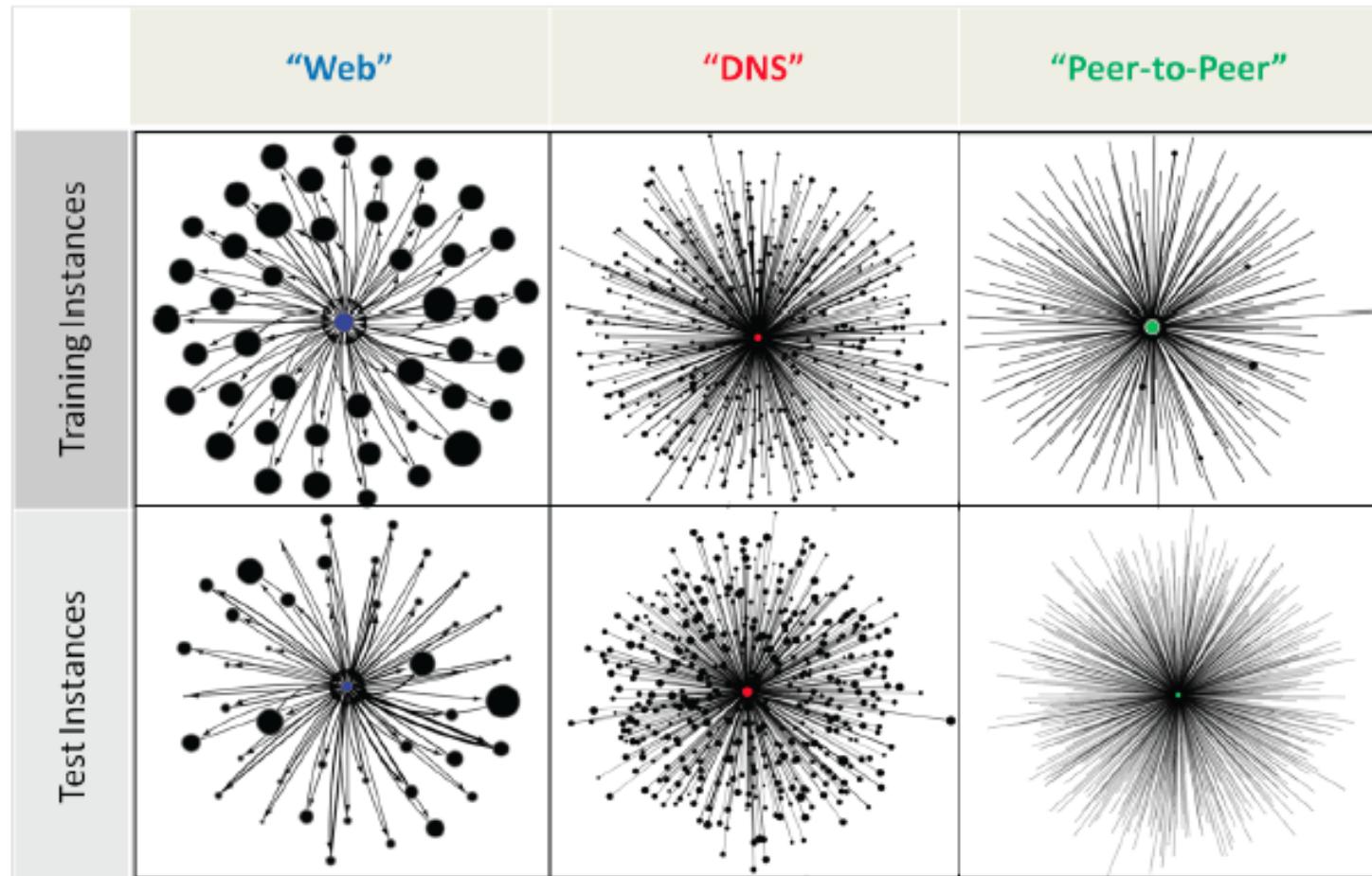


ReFeX: Structural Features

- **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

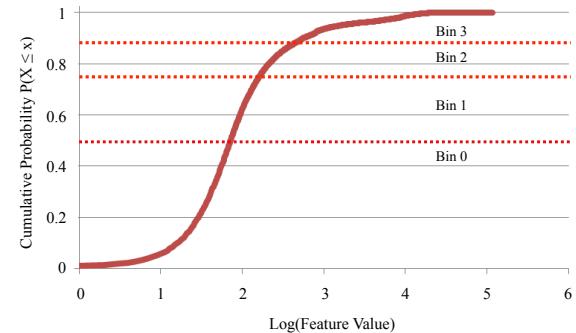


Node sizes indicate communication volume relative to the central node in each frame.



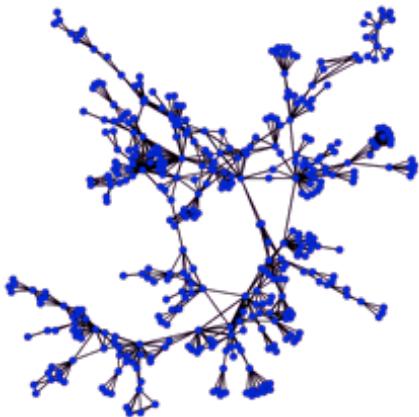
ReFeX (continued)

- 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



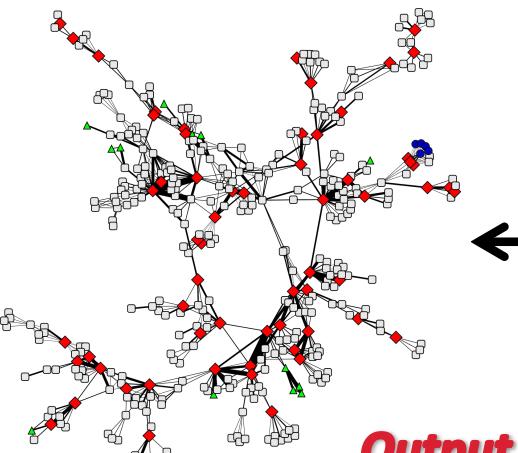
Role Extraction

Input



→ **Recursively extract features**

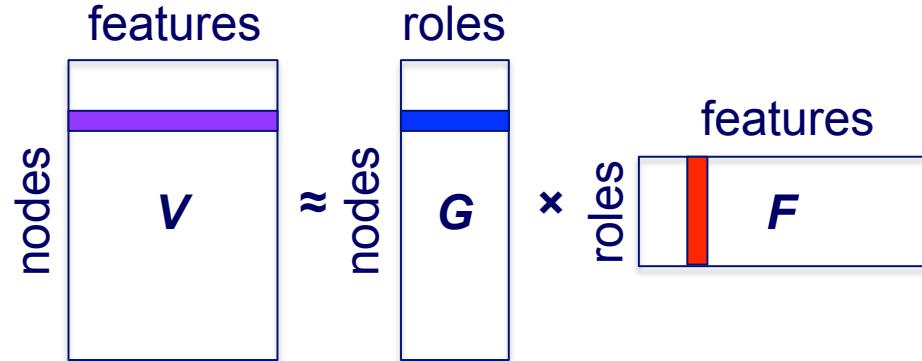
Nodes → Features

Nodes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	629	630	631	632	633	634	635	636	637	638	639	640	641	642	643	644	645	646	647	648	649	650	651	652	653	654	655	656	657	658	659	660	661	662	663	664	665	666	667	668	669	660	661	662	663	664	665	666	667	668	669	670	671	672	673	674	675	676	677	678	679	680	681	682	683	684	685	686	687	688	689	690	691	692	693	694	695	696	697	698	699	700	701	702	703	704	705	706	707	708	709	710	711	712	713	714	715	716	717	718	719	720	721	722	723	724	725	726	727	728	729	730	731	732	733	734	735	736	737	738	739	740	741	742	743	744	745	746	747	748	749	750	751	752	753	754	755	756	757	758	759	760	761	762	763	764	765	766	767	768	769	770	771	772	773	774	775	776	777	778	779	770	771	772	773	774	775	776	777	778	779	780	781	782	783	784	785	786	787	788	789	790	791	792	793	794	795	796	797	798	799	800	801	802	803	804	805	806	807	808	809	810	811	812	813	814	815	816	817	818	819	820	821	822	823	824	825	826	827	828	829	830	831	832	833	834	835	836	837	838	839	840	841	842	843	844	845	846	847	848	849	850	851	852	853	854	855	856	857	858	859	860	861	862	863	864	865	866	867	868	869	870	871	872	873	874	875	876	877	878	879	880	881	882	883	884	885	886	887	888	889	880	881	882	883	884	885	886	887	888	889	890	891	892	893	894	895	896	897	898	899	900	901	902	903	904	905	906	907	908	909	910	911	912	913	914	915	916	917	918	919	920	921	922	923	924	925	926	927	928	929	930	931	932	933	934	935	936	937	938	939	940	941	942	943	944	945	946	947	948	949	950	951	952	953	954	955	956	957	958	959	960	961	962	963	964	965	966	967	968	969	970	971	972	973	974	975	976	977	978	979	980	981	982	983	984	985	986	987	988	989	990	991	992	993	994	995	996	997	998	999	1000


D. Koutra & T. Eliassi-Rad & C. Faloutsos

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 $V \approx GF$



	Features	
Nodes	1	2
1411	1	2
1410	1	1
338	0	0
339	0	0
1412	1	1
941	0	0
1414	1	1
942	1	1
1415	1	1
1416	0	0
940	0	1
1419	0	1
945	1	1
1417	0	1
1418	1	1
946	1	1
33	0	0
1412	1	1
1417	1	1
30	1	2
1416	1	1
944	1	1
331	1	2
943	1	1
336	0	0
337	1	1
947	1	1
338	0	0
339	0	0
340	0	0
341	0	0
342	0	0
343	0	0
344	0	0
345	0	0
346	0	0
347	0	0
348	0	0
349	0	0
350	0	0
351	0	0

- RolX 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
- 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 Lloyd-Max quantization with Huffman codes
 - Errors in $V-GF$ are not distributed normally, RolX uses KL divergence to compute E

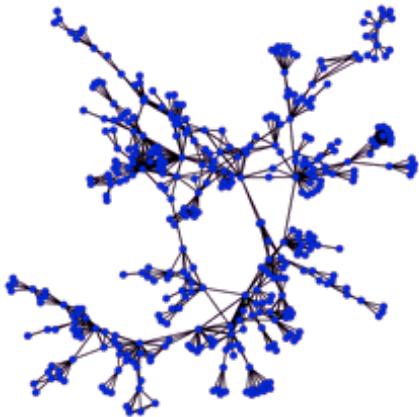
$$M = \bar{b}r(n + f)$$

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



Role Extraction

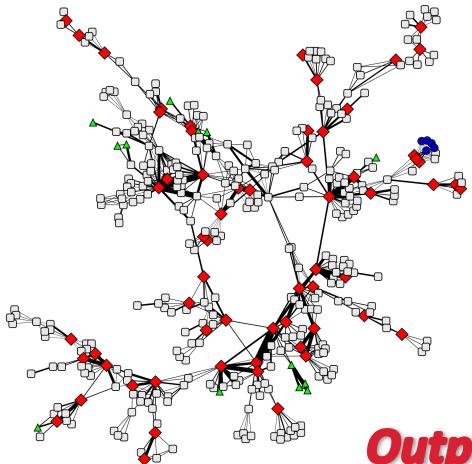
Input



→ **Recursively extract features**

Nodes	Features																	
1411	0	1	0	1	0	0	0	1	1	0	0	1	1	1	2	2	2	
1410	0	1	0	1	0	1	0	0	1	0	0	1	0	1	1	1	1	1
338	0	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0
339	1	0	0	0	0	2	0	1	0	0	2	0	0	1	0	0	0	0
1415	0	1	0	1	0	1	0	0	0	0	1	0	0	1	1	1	1	1
341	0	1	0	1	0	1	0	0	0	0	1	0	0	1	1	1	1	1
3414	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1	1	1
342	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
3413	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	1	1	1
1412	0	0	0	0	0	0	0	0	1	2	0	1	1	0	0	1	2	0
3419	0	0	1	0	0	0	1	0	1	1	0	1	1	1	1	1	1	1
345	0	1	0	0	0	0	0	0	0	2	0	1	0	0	2	1	3	1
3418	0	0	1	0	0	0	1	0	1	0	0	0	0	0	2	0	1	1
346	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	1	1	2
333	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0
1417	0	1	0	1	0	2	0	0	1	0	1	0	1	0	1	1	1	1
343	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
330	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1416	0	1	0	0	1	0	2	0	0	2	0	0	0	0	0	2	0	0
344	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
331	0	0	1	0	0	1	0	0	0	0	2	0	0	0	0	0	0	0
349	0	0	0	0	0	2	0	0	1	0	0	1	0	0	0	0	0	0
336	0	0	0	0	0	0	2	0	0	1	1	1	1	0	0	0	1	0
337	1	1	0	0	0	1	2	0	0	1	1	1	1	0	1	1	1	1
347	1	0	0	0	0	2	0	1	0	0	2	0	1	0	1	0	0	0
334	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
335	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
331	1	0	0	0	0	1	0	0	0	0	2	0	0	0	0	2	0	0

→ **Automatically factorize roles**



Output



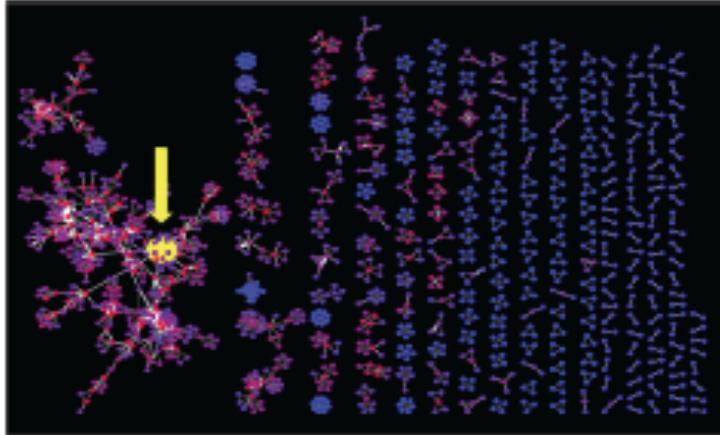
Experiments on Role Discovery

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

Details in Henderson *et al.* KDD 2012

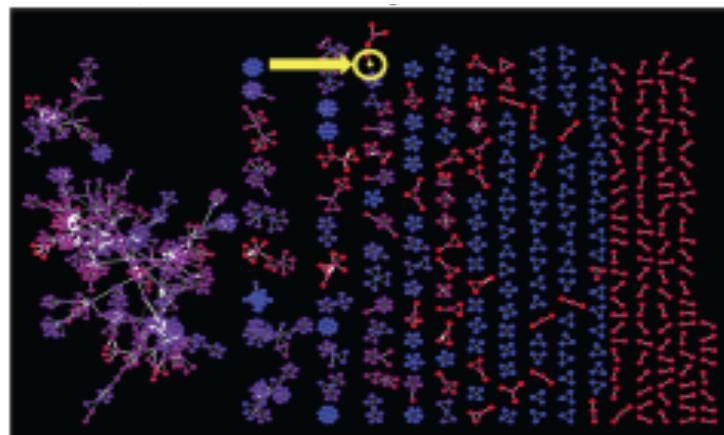


Role Query

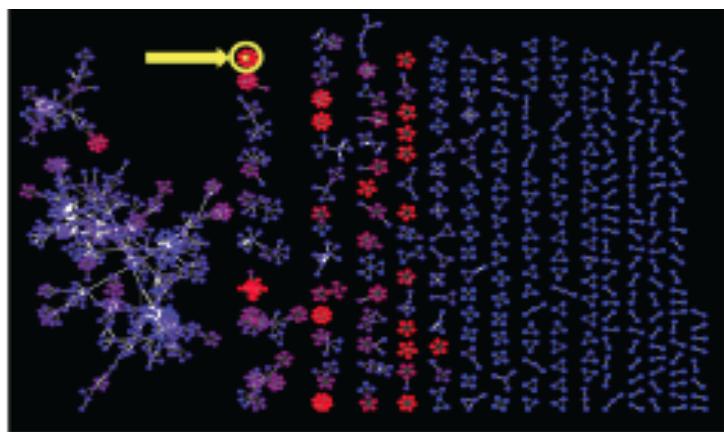


Node Similarity for M.E.J. Newman
(*bridge*)

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



Node Similarity for J. Rinzel (*isolated*)

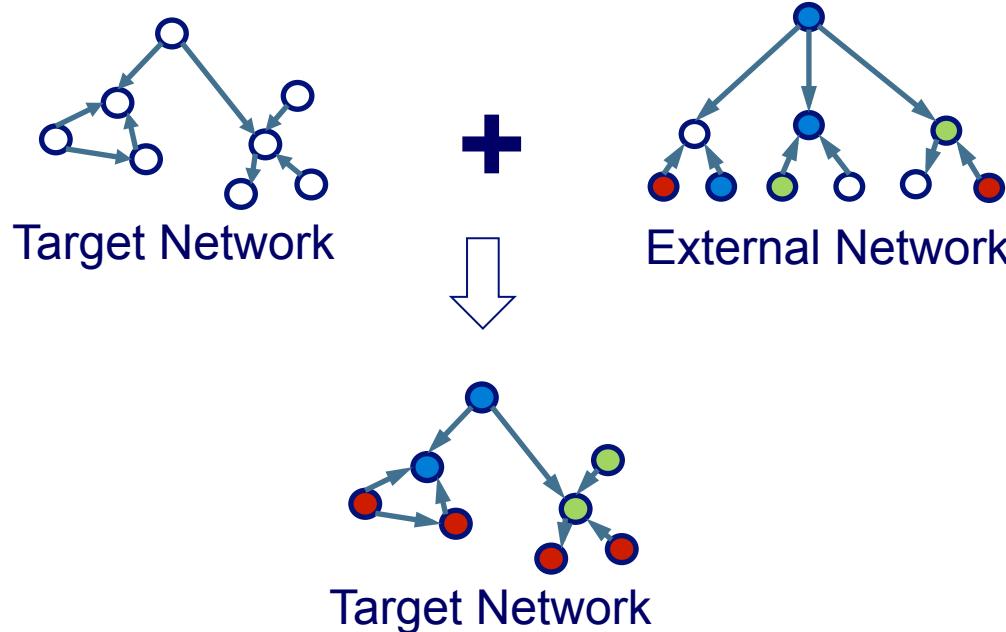


Node Similarity for F. Robert (*cliquey*)



Role Transfer

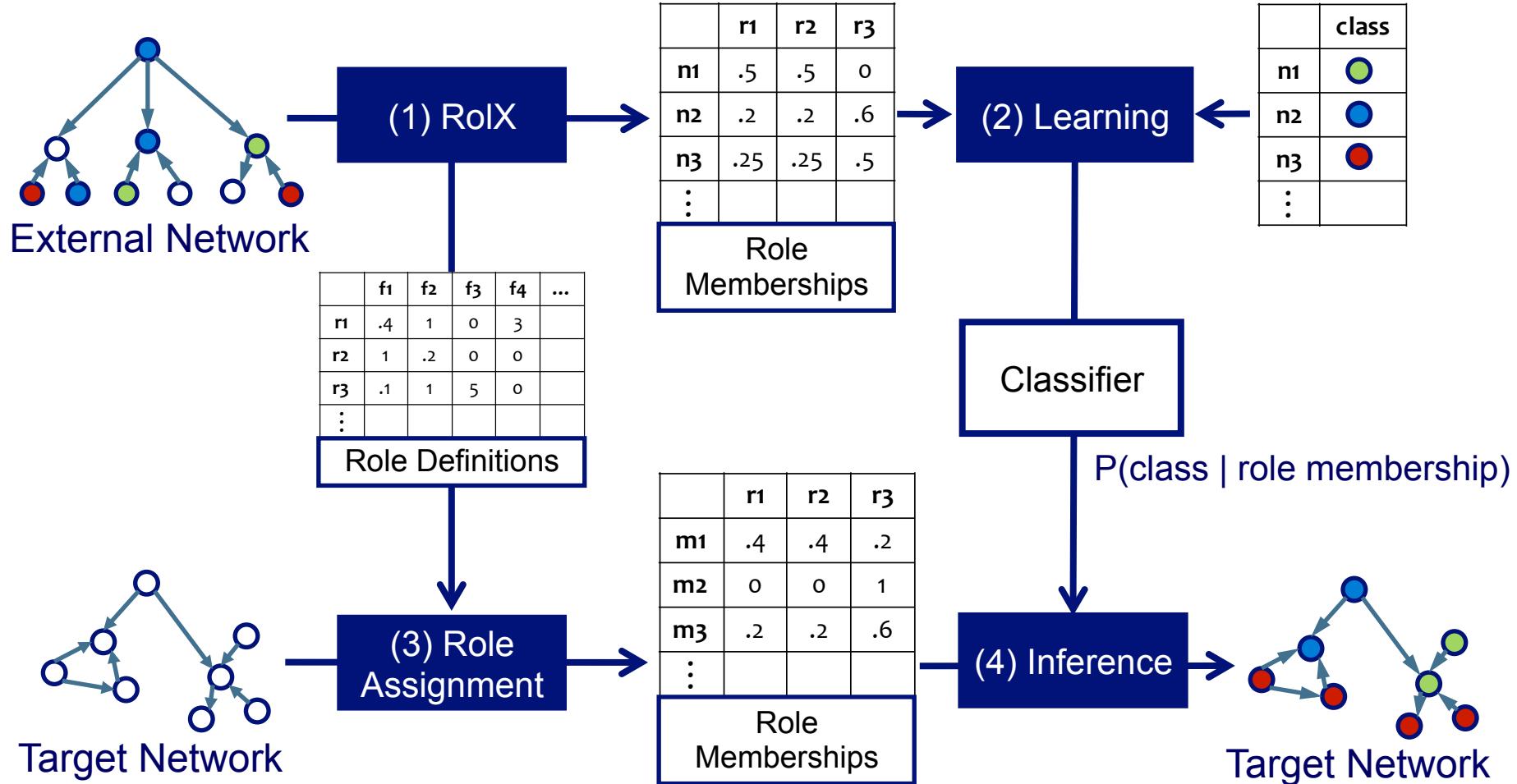
- Question: How can we use labels from an external source to predict labels on a network with no labels?



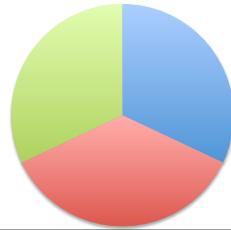
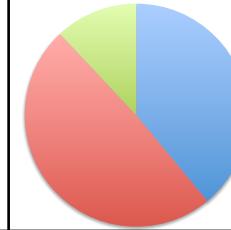
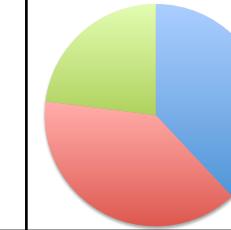
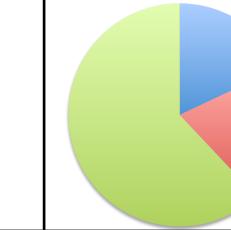
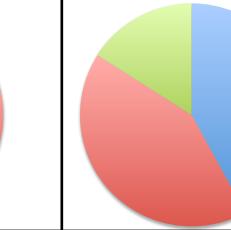
- Conjecture: Nodes with similar roles are likely to have similar labels



Role Transfer = RolX + SL

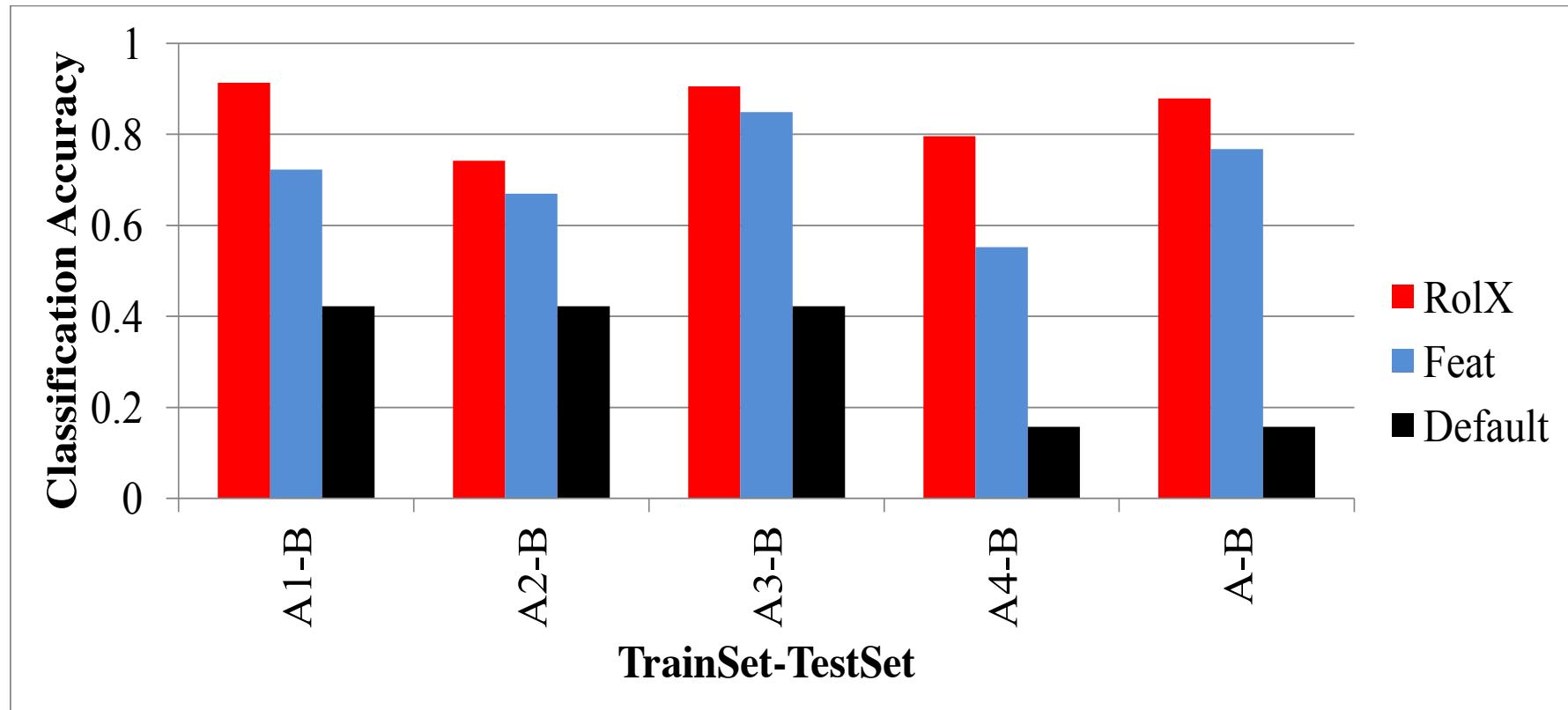


Data for Role Transfer

	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 Distribution					
	 Web	 DNS	 P2P		



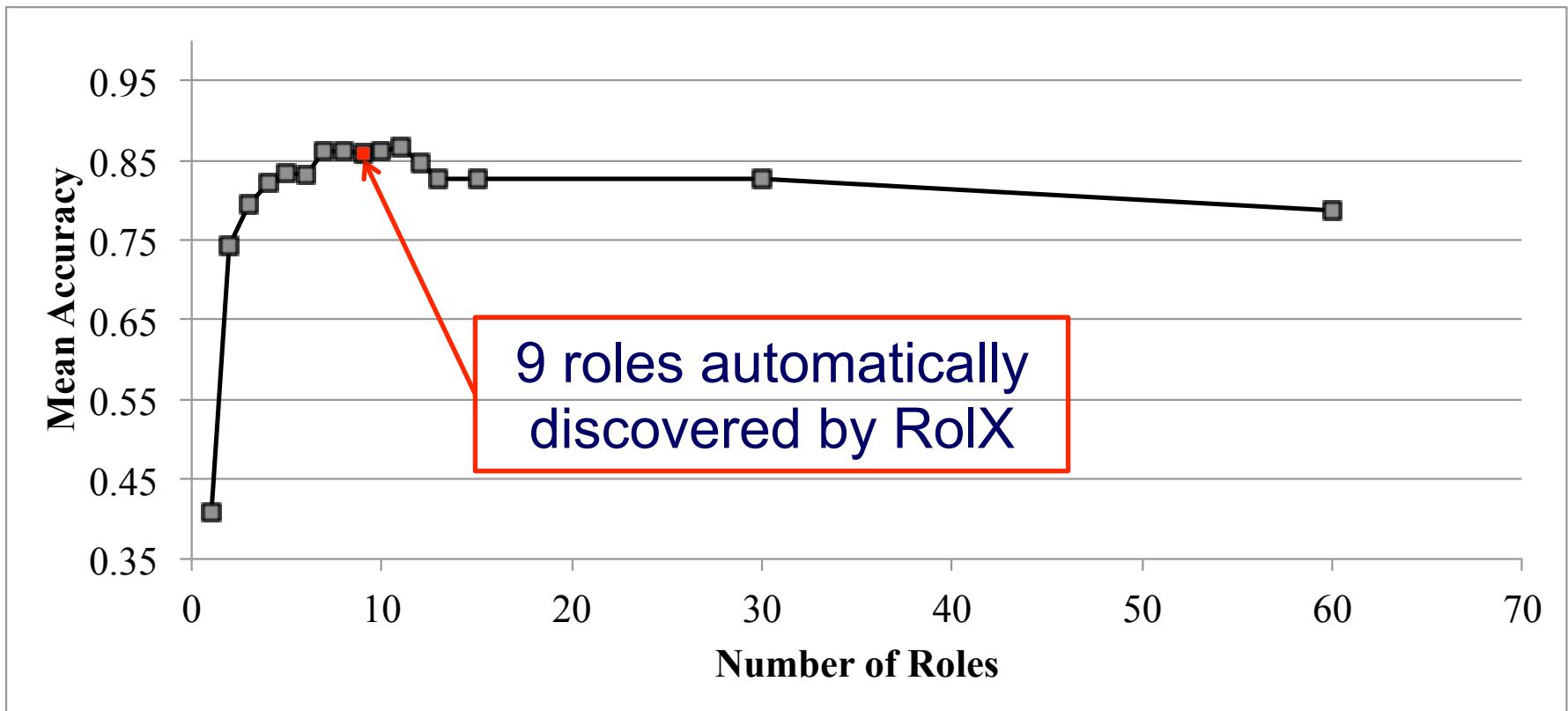
Role Transfer Results



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



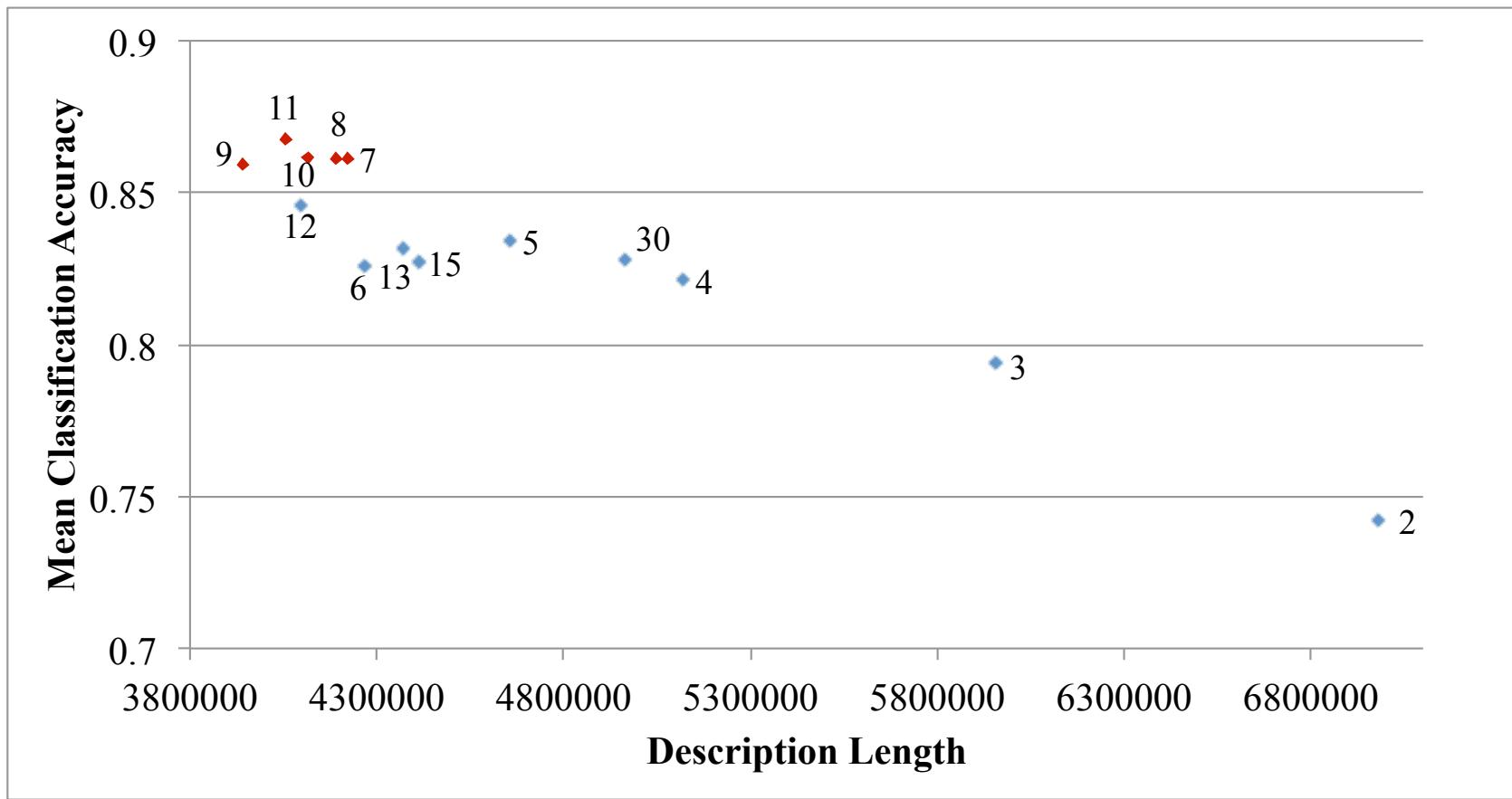
Model Selection



RolX selects high accuracy model sizes



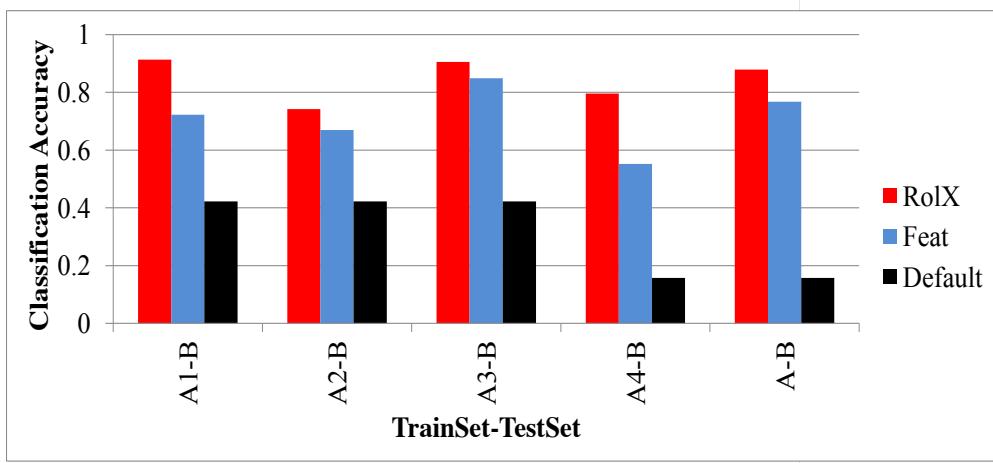
Model Selection (continued)



Classification accuracy is highest when RoIX selection criterion is minimized

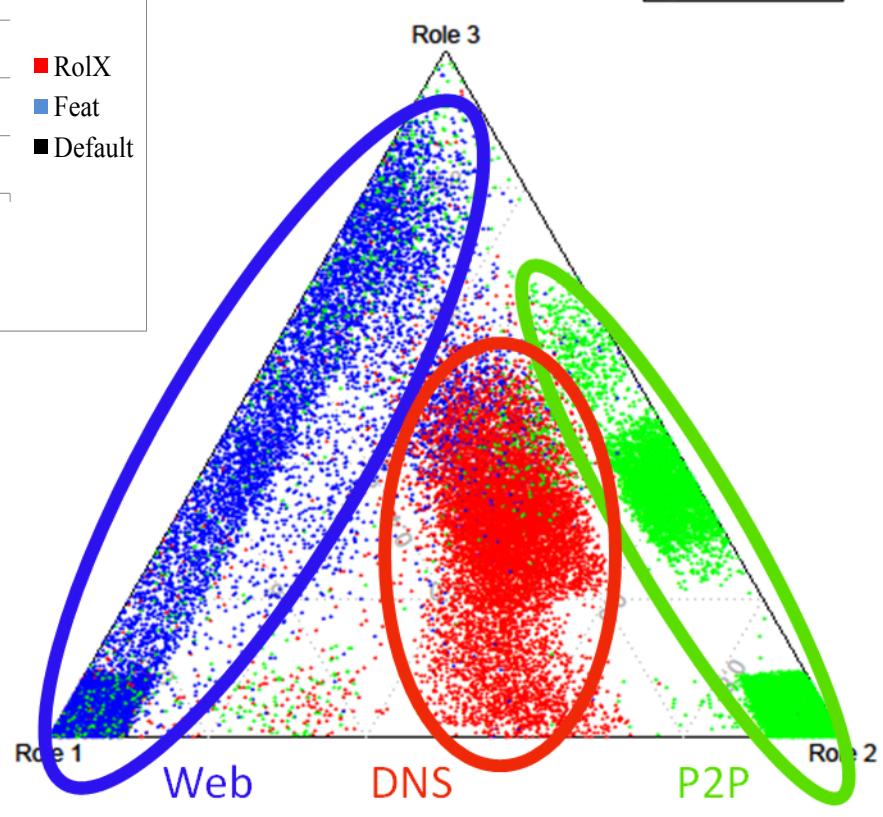


Role Space



Applications x Roles

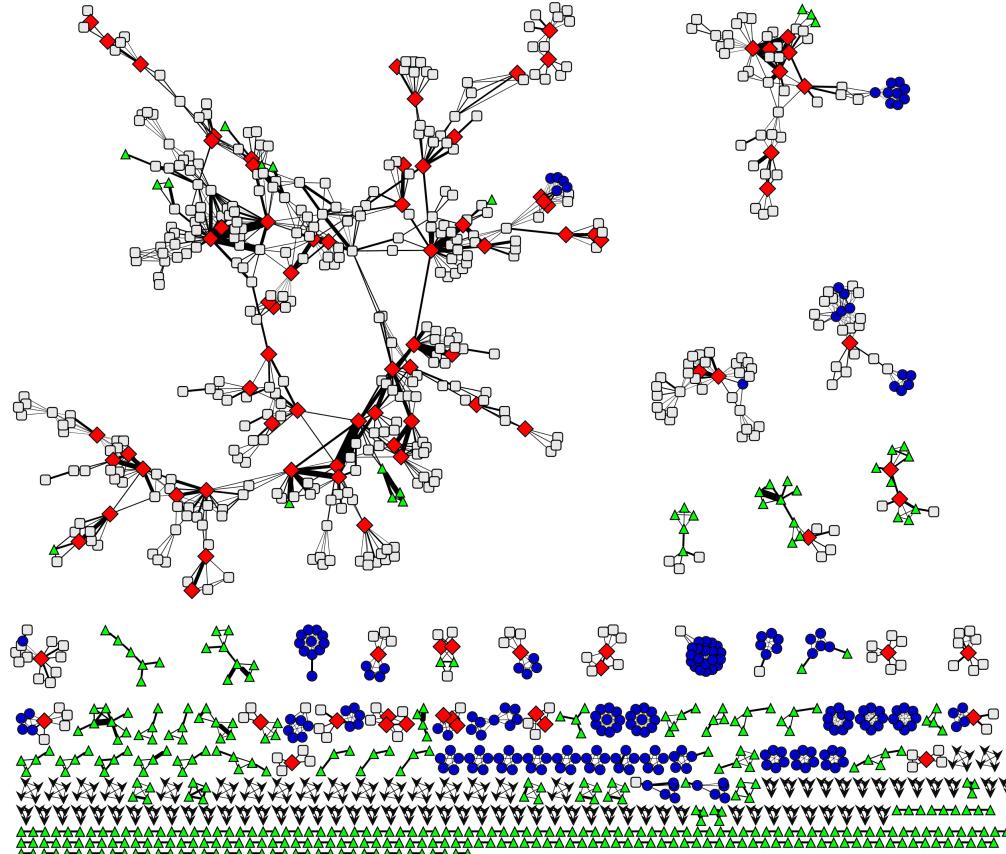
Applications
● DNS
● P2P
● Web



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



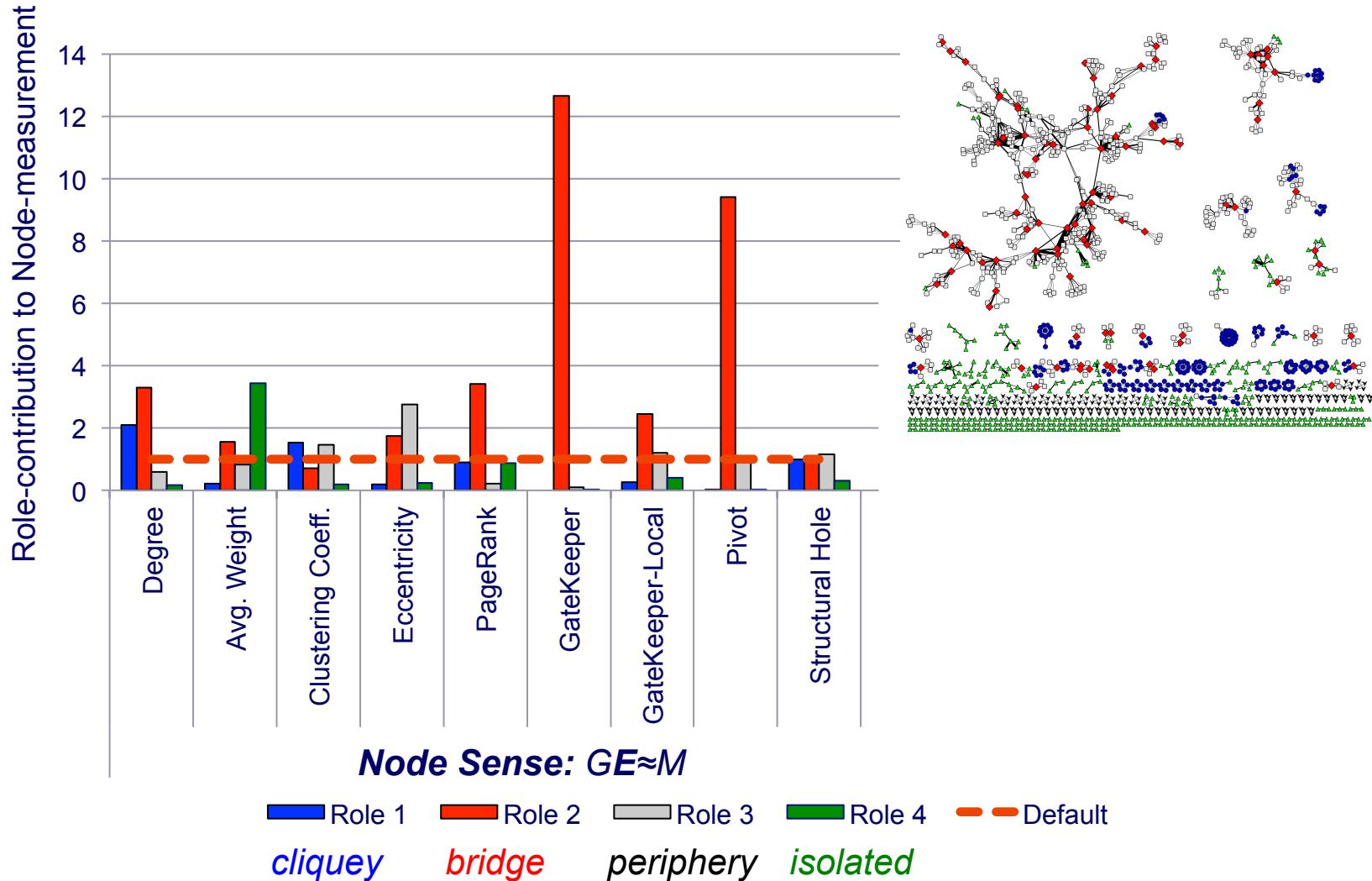
Automatically Discovered Roles



Network Science Co-authorship Graph
[Newman 2006]

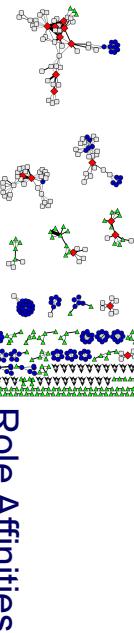
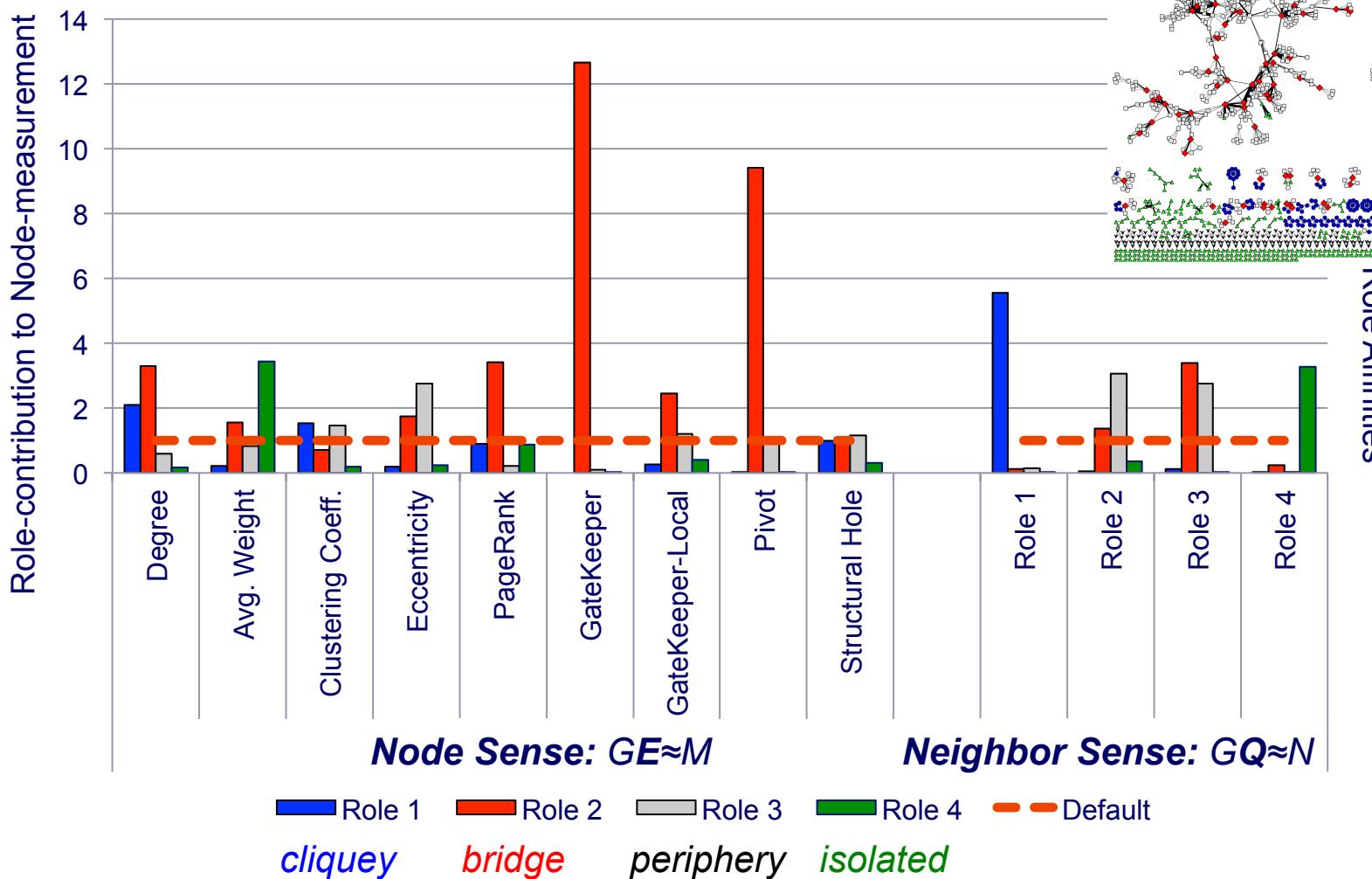


Making Sense of Roles



Making Sense of Roles

Roles can be interpreted using topological measures & role homophily



Role Affinities

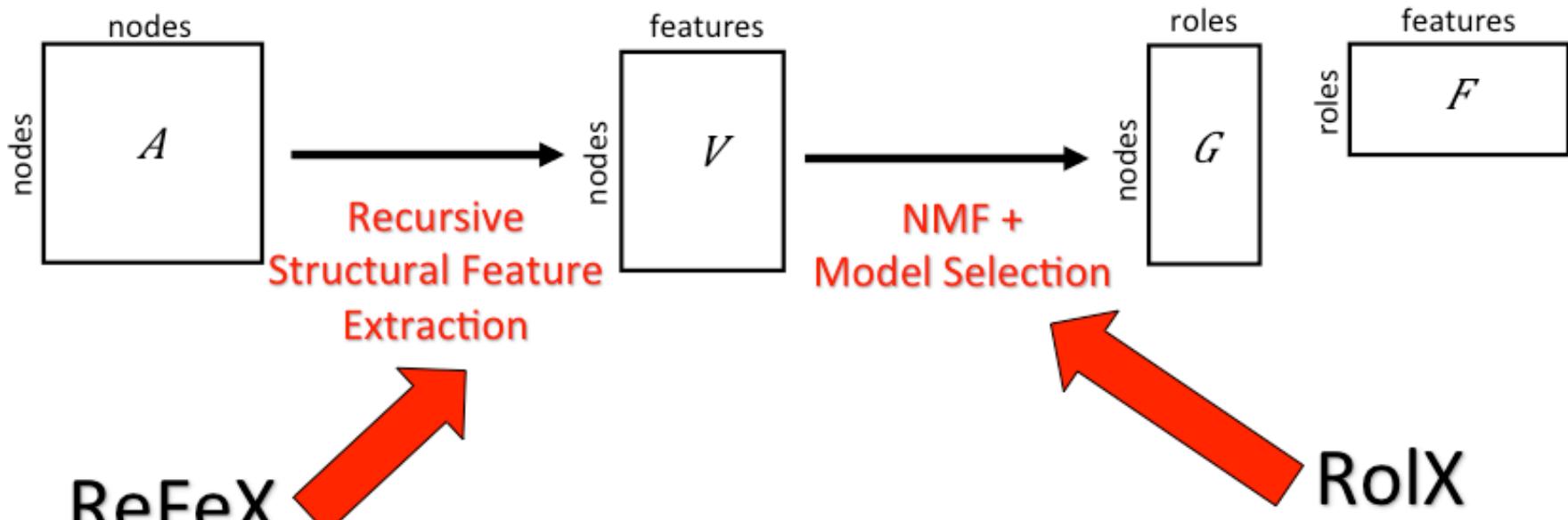


GLRD: Guided Learning for Role Discovery

- 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

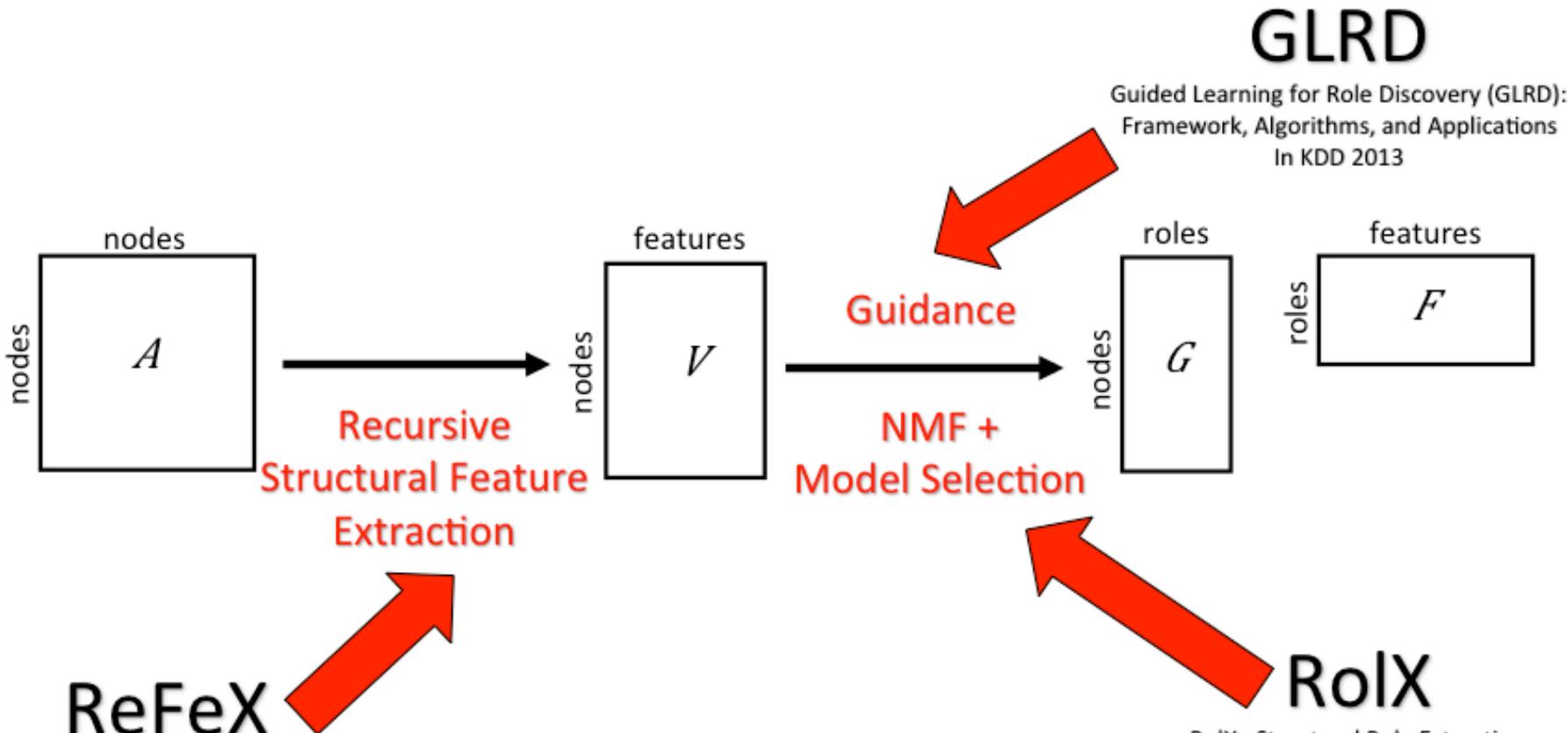


It's Who You Know: Graph Mining Using
Recursive Structural Features
In KDD 2011

RoIX: Structural Role Extraction
& Mining in Large Graphs
In KDD 2012.



GLRD



It's Who You Know: Graph Mining Using
Recursive Structural Features
In KDD 2011

Guided Learning for Role Discovery (GLRD):
Framework, Algorithms, and Applications
In KDD 2013



GLRD Framework

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

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

$$\text{subject to} \quad g_i(\mathbf{G}) \leq d_{Gi}, \quad i = 1, \dots, t_G$$

$$f_i(\mathbf{F}) \leq d_{Fi}, \quad i = 1, \dots, t_F$$

where g_i and f_i are convex functions.

- Use an alternative least squares (ALS) formulation
 - Do not alternate between solving for the entire \mathbf{G} and \mathbf{F}
 - Solve for one column of \mathbf{G} or one row of \mathbf{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

$$\operatorname{argmin}_{\mathbf{G}, \mathbf{F}} \quad \|\mathbf{V} - \mathbf{GF}\|_2$$

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

$$\forall i \quad \|\mathbf{G}_{\bullet i}\|_1 \leq \epsilon_G$$

$$\forall i \quad \|\mathbf{F}_{i \bullet}\|_1 \leq \epsilon_F$$

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

$$\operatorname{argmin}_{\mathbf{G}, \mathbf{F}} \|\mathbf{V} - \mathbf{GF}\|_2$$

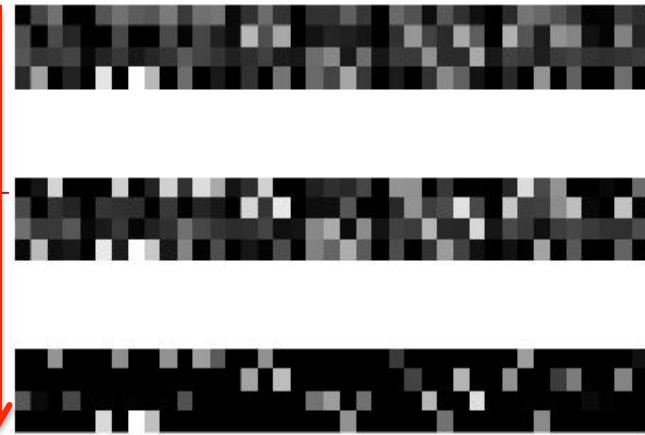
subject to: $\mathbf{G} \geq 0, \mathbf{F} \geq 0$

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

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

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

more diverse



Diverse Roles and Sparse Roles

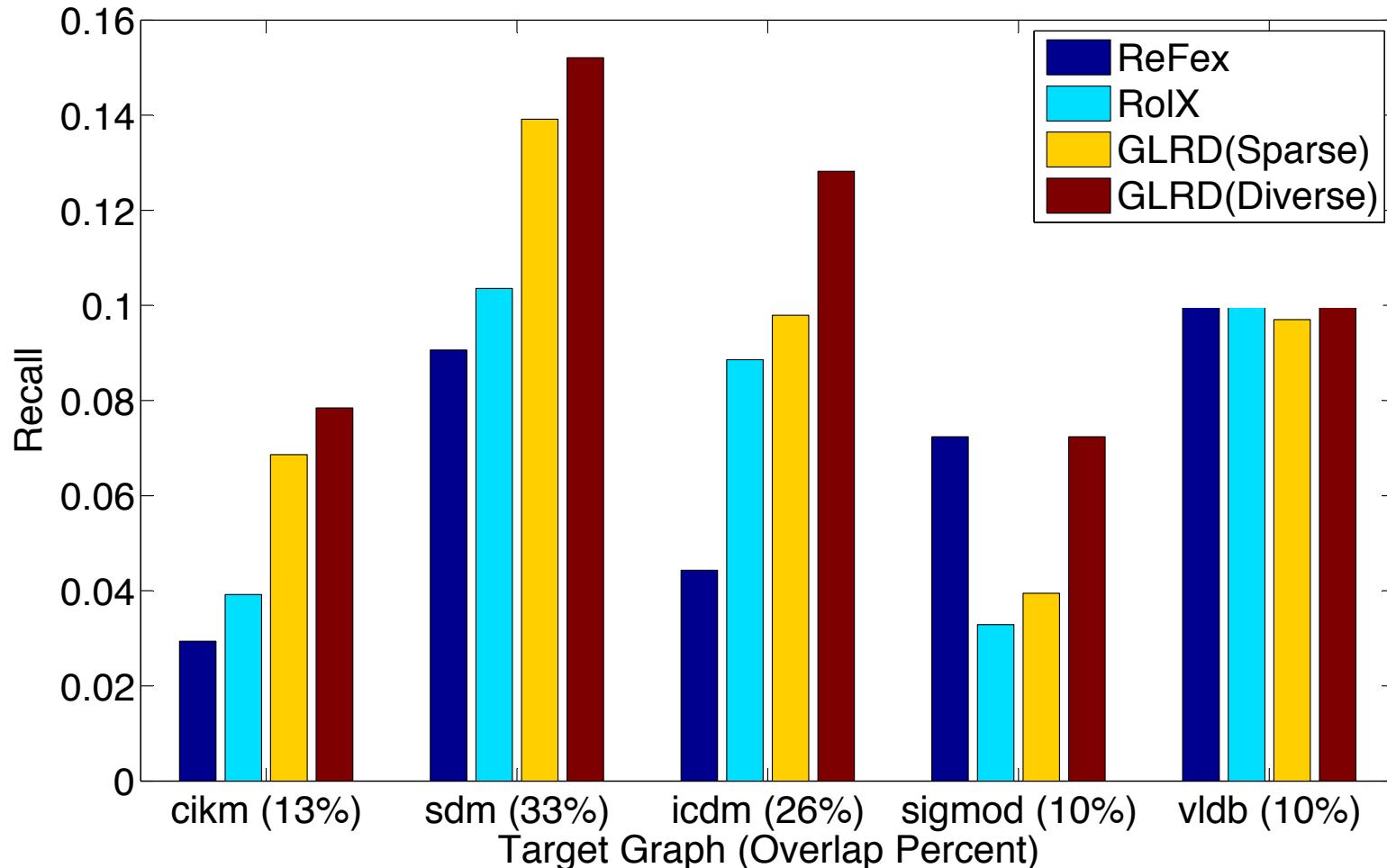
- Question: Can diversity and sparsity constraints create better role definitions?
- Conjecture: Better role definitions will better facilitate other problems such as identity resolution across graphs
- Experiment: Compare graph mining results using various methods for role discovery

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



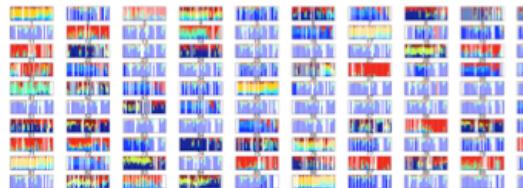
Identity Resolution across Networks



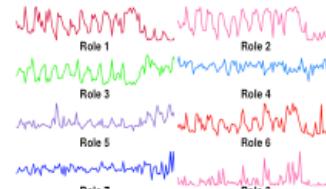
Modeling Dynamic Graphs with Roles

- Introduced by Rossi et al. [WSDM 2013]

1. Identify dynamic patterns in node behavior

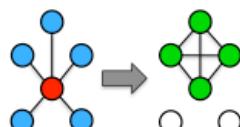


Evolving mixed-role memberships



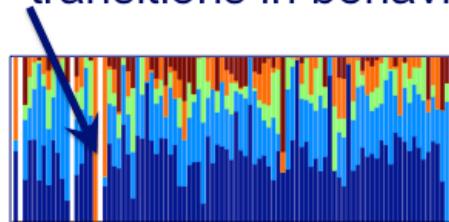
Role contributions

2. Predict future structural changes



Transition from star to clique

3. Detect unusual transitions in behavior



- Given G_{t-1} and G_t find a transition model T that minimizes the functional:
- All models predict G_{t+1} using G_t as

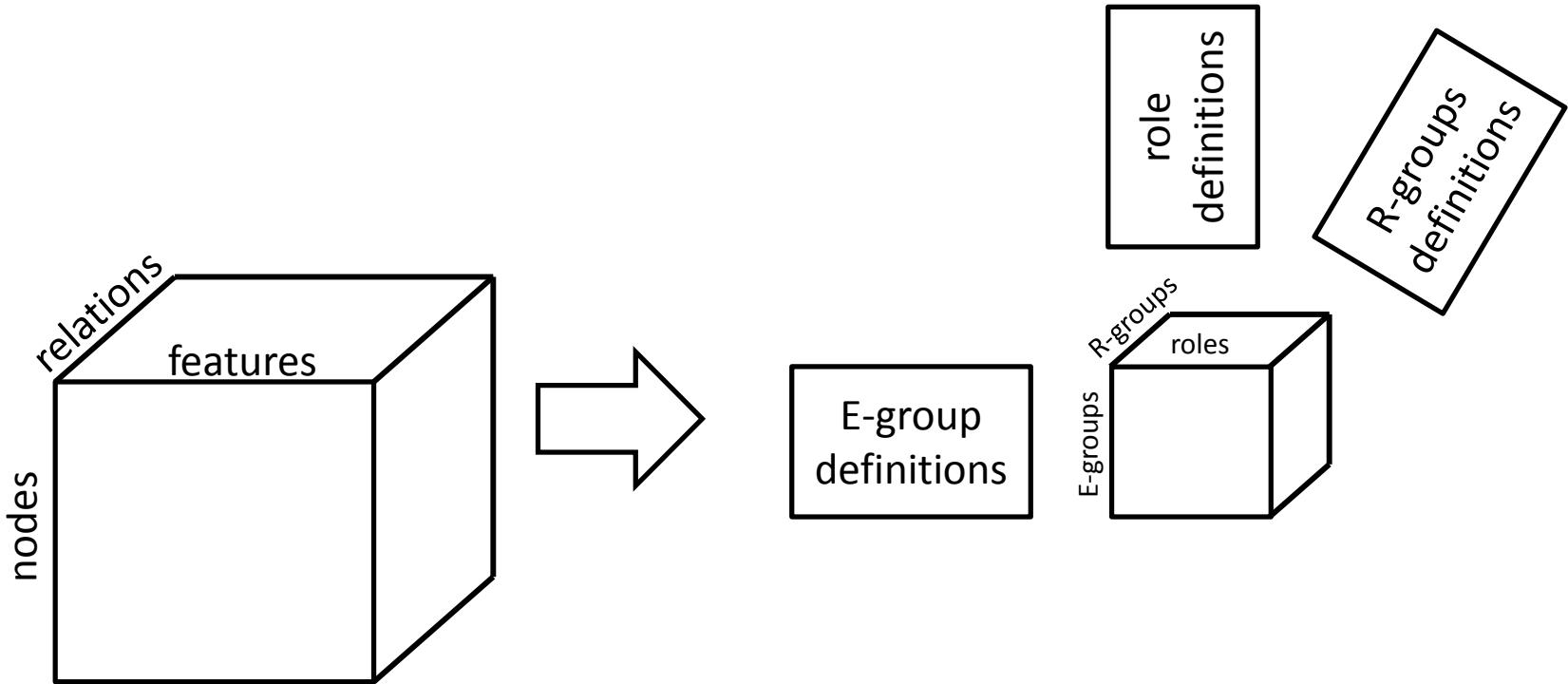
$$f(G_t, G_{t-1}) = \frac{1}{2} \|G_t - G_{t-1}T\|_F^2$$

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



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



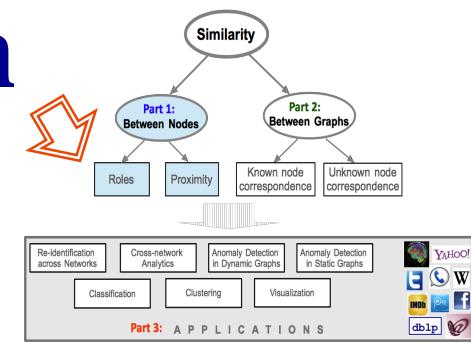
Roadmap

- 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 Berlingario, 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. Zelenzy, 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

