

Smart Roles: Inferring Professional Roles in Email Networks Di Jin₁* Mark Heimann₁* Tara Safavi₁ Mengdi Wang₂ Wei Lee₃ Lindsay Snider₃ and Danai Koutra, ¹University of Michigan, Ann Arbor ²University of Pittsburgh ³Trove AI

Motivation: bring order to flooded email inboxes - *Prioritize* emails from important senders - *Recommend* useful connections **Goal**: Infer professional roles of email users

Intuition: Professional role inference ≈ structural role inference in email networks **Method:** Design *node embedding* method, EMBER, to efficiently capture structural roles

S1 :	Capture	each	user's	s local	structure
-------------	---------	------	--------	---------	-----------



Email communication is	We do
Higher-order (indirect connections matter)	Define structural bell histograms based of communication up to $\mathbf{b}_{u}^{+} = \sum_{k=0}^{K} \delta^{k} \mathbf{b}_{u}^{k}$
Weighted by # of emails two users exchange	Make histograms bin weights between a uncertainty connections $b_{u,d}^{k+} = \sum_{v \in D_u^{k+}} \text{path_weight}$
Directed from sender to receiver	Create, concatenate histograms from pat outgoing and income $\mathbf{b}_u = [\mathbf{b}_u^+, \mathbf{b}_u^-]$

S2: *Embed* users by comparing their local structure to a small number of *landmark users*







havior n paths of o length K

n *path* user and her

ht $\left(\mathcal{P}_{u
ightarrow v}^{k+}
ight)$

separate ihs following *ing* edges

New email network datasets collect from Trove AI

- Multiple companies
- Varying company sizes

Group users' job titles into broad catego Officers, middle management, and w Infer role of employees who do not provi

Effective: quantitative improvement in classification accuracy

	State-of-	the-art	baselir	nes rep		several fa	milies of t		Jes Granhwave	Unweigh	ted/undire	cted var
	JINA	KOIX	LIIDI	LINL	Бесриак	noue2vee	struczvec	DIVOR	Oraphiwave	LIVIDLIC-O	LWIDLK-D	LINIDLI
Frove-318	.7605	.5670	.6908	.6618	.7602	.7648	.7799	.7131	.7685	.7749	.7563	.7625
Frove-183	.7648	.5787	.7718	.5657	.8071	.8223	.8264	4925	.6391	.7986	.7838	.8186
rove-141	.6738	.5591	.7409	.7102	.7191	.7474	.7391	.6235	.7112	.7291	.7309	.6971
Frove-98	.6676	.5177	.6323	.6872	.5587	.6198	.6498	.5329	$.7177^{*}$.6040	.5857	.6333
Frove-19	.5429	.6981	.6248	.7184	.5531	.5959	.6102	.6089	.7157	.6837	.7204	.6939
「rove−2K	.6305	.5212	.6622	.6771	.6769	.6780	.6802	.6527	.6594	.6689	.6345	.6677
Frove	.6633	.5280	5454	—	.6866	.6951				.6905	.7141	.7122
Enron	.6205	.5197	.5000	.6931	.7201	.7389		.5709		.7393	.7347	.7305

Practical: scales to *millions* of users

EMBER	2.50	16.87	830.80	84.98	unerent size companies	Mapping professio	profe onal ro	い こ つ
Graphwave	2.73	5.66	>12h	>12h	different size companies		Officer M	1ç
DNGR	21.05	72.83	>12h	>12h	Manning roles across	Irove-19	0.75	J
struc2vec	17.48	188.65	>12h	29286.38	Trove-98	T 10	0.75	0
node2vec	2.85	24.55	3484.05	254.60	OfficerMgmt.Worker	Irove-98	0.57	J
DeepWalk	3.12	21.59	2464.13	255.84		T 00		
LINE	171.95	153.12	>12h	267.48	Worker - 0.30 0.56 0.14	Irove-141	0.17	
LinBP	0.54	2.88	14607.44	1038.09	₽ E	T	0.17	
RolX	0.14	0.16	2150.53	205.92	Mgmt 0.35 0.51 0.15	Irove-183	0.24	
SNA	6.32	16.45	3193.26	333.33	33	Travia 102	0.24	
	Trove-318	Trove-2K	Trove	Enron	Officer - 0.33 0.58 0.08	Trove-318 ⁻	0.13	0



Output: network with inferred roles

2 Multi-class classification

2-dim embedding space

ted		Employees	Connections	Email exchanges	# Offi mgmt
	Trove-19	19	47	274	4
	Trove-98	98	101	1769	53
	Trove-141	141	1 242	9565	23
	Trove-183	183	3136	21 655	16
	Trove-318	318	1026	12 643	30
nes:	Trove-2K	2414	16 281	183 443	495
workers	Trove	9 989 507	40 290 044	568 678 419	495
ide titles	Enron	75 416	319 935	2 064 442	31

Insignttul : can anal	iyze role <i>comparability</i>
------------------------------	--------------------------------



Motivation: bring order to flooded email inboxes - *Prioritize* emails from important senders - *Recommend* useful connections

Goal: Infer professional roles of email users

Intuition: Professional role inference ≈ structural role inference in email networks **Method:** Design *node embedding* method, EMBER, to *efficiently* capture structural roles

S1: Capture each user's local structure

Email is:



Higher-order (indirect connections matter)

- Form structural behavior histograms based on patterns of length \leq K paths of communication
- Weighted by # of emails two users exchange
 - Bin path weights between a user and connections
- **Directed** from sender to receiver
 - Create. concatenate separate histograms from paths following outgoing and incoming edges $\mathbf{b}_u = [\mathbf{b}_u^+, \mathbf{b}_u^-]$

S2: Embed local structure

User-to-*landmark* similarities

- Sample landmarks ~ degree
- Structural similarity between users:

$$sim(u,v) = e^{-||\mathbf{b}_u - \mathbf{b}_v||}$$

- Approximate decomposition of pairwise user structural similarity matrix
- Can embed only important subsets of users

From SVD of psuedoinverse of pairwise landmark similarities



$$\mathbf{Y} = \mathbf{C}\mathbf{U}\mathbf{\Sigma}$$



Smart Roles: Inferring Professional Roles in Email Networks Di Jin¹* Mark Heimann¹* Tara Safavi¹ Mengdi Wang² Wei Lee³ Lindsay Snider³ and Danai

¹University of Michigan, Ann Arbor

²University of Pittsburgh ³Trove AI





Output: network with inferred roles

 $\mathbf{b}_{u}^{+} = \sum_{k=0}^{K} \delta^{k} \mathbf{b}_{u}^{k+}$

 $b_{u,d}^{k+} = \sum_{u \to v} \operatorname{path_weight} \left(\mathcal{P}_{u \to v}^{k+} \right)$

1/2

New email network datasets collect from Trove AI

- Multiple companies
- Varying company sizes

Group users' job titles into broad catego Officers, middle management, and w Infer role of employees who do not provi

Effective: quantitative improvement in classification accuracy State-of-the-art baselines, several families of techniques

	SNA	RolX	LinBP	LINE	DeepWalk	node2vec	struc2vec	DNGR	Graphwave	EMBER-U	EMBER-D	EMBER-V
rove-318	.7605	.5670	.6908	.6618	.7602	.7648	.7799	.7131	.7685	.7749	.7563	.7625
rove-183	.7648	.5787	.7718	.5657	.8071	.8223	.8264	4925	.6391	.7986	.7838	.8186
rove-141	.6738	.5591	.7409	.7102	.7191	.7474	.7391	.6235	.7112	.7291	.7309	.6971
rove-98	.6676	.5177	.6323	.6872	.5587	.6198	.6498	.5329	.7177*	.6040	.5857	.6333
rove-19	.5429	.6981	.6248	.7184	.5531	.5959	.6102	.6089	.7157	.6837	.7204	.6939
rove-2K	.6305	.5212	.6622	.6771	.6769	.6780	.6802	.6527	.6594	.6689	.6345	.6677
rove	.6633	.5280	5454	—	.6866	.6951				.6905	.7141	.7122
nron	.6205	.5197	.5000	.6931	.7201	.7389		.5709		.7393	.7347	.7305

Practical: scales to *millions* of users

EMBER	2.50	16.87	830.80	84.98		u				πράπου	Mapping		fes
Graphwave	2.73	5.66	>12h	>12h		ivi di	appill fforon	y iui nt ciza	5 a	nnaniae		Officer	Μ
DNGR	21.05	72.83	>12h	>12h		Л Л	annin	a rol	00.00	ornee	Irove-19	0.75	0
struc2vec	17.48	188.65	>12h	29286.38					Trove-	98	T 10	0.75	0
node2vec	2.85	24.55	3484.05	254.60				Office	rMgmt	.Worker	Irove-98	0.57	0
DeepWalk	3.12	21.59	2464.13	255.84				- 1	1		T 00		0
LINE	171.95	153.12	>12h	267.48		۷	Vorker -	0.30	0.56	0.14	Irove-141	0.17	U
LinBP	0.54	2.88	14607.44	1038.09	1	Tro L					Trovia 141	0 1 7	\cap
RolX	0.14	0.16	2150.53	205.92		Ve-	Mgmt	0.35	0.51	0.15	110ve-183	0.24	
SNA	6.32	16.45	3193.26	333.33		318					$T_{rov} = 102$	0.24	0
	Trove-318	Trove-2K	Trove	Enron		(Officer -	0.33	0.58	0.08	Trove-318	0.13	0



(2) Multi-class classification

2-dim embedding space

ted		Employees	Connections	Email exchanges	# Offi mgmt
	Trove-19	19	47	274	4
	Trove-98	98	101	1769	53
	Trove-141	141	1 242	9565	23
	Trove-183	183	3136	21 655	16
ricou	Trove-318	318	1026	12 643	30
mes.	Trove-2K	2414	16 281	183 443	495
workers	Trove	9 989 507	40 290 044	568 678 419	495
vide titles	Enron	75 416	319 935	2 064 442	31

Unweighted/undirected variants

Insightful: can	analyze role	comparability
-----------------	--------------	---------------