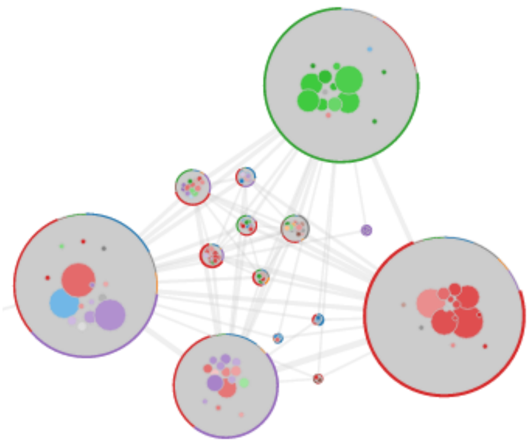
CS 512 Data Mining Principles

Network as a Context

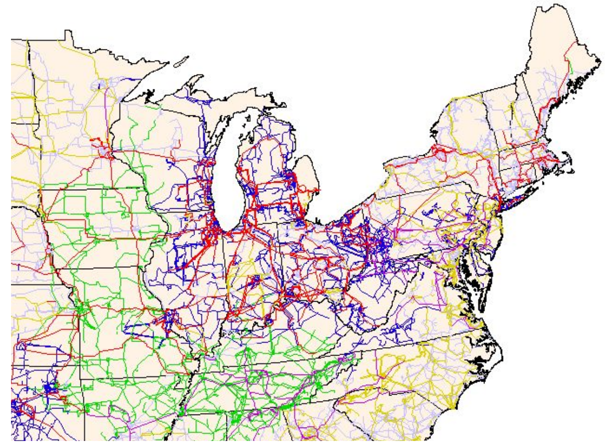
Hanghang Tong, Computer Science, Univ. Illinois at Urbana-Champaign, 2021

Suggested studying time: 4/16/2021-4/22/2021

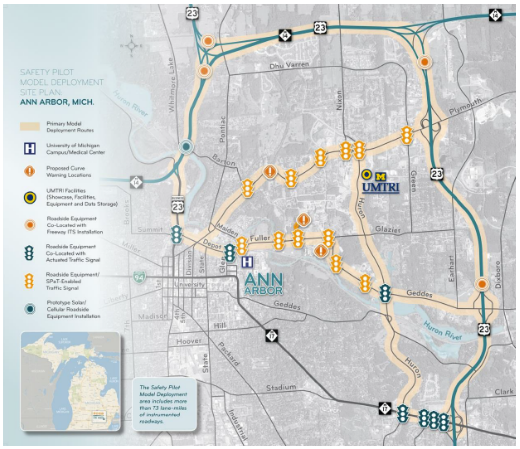
Networks & Graphs Are Everywhere!



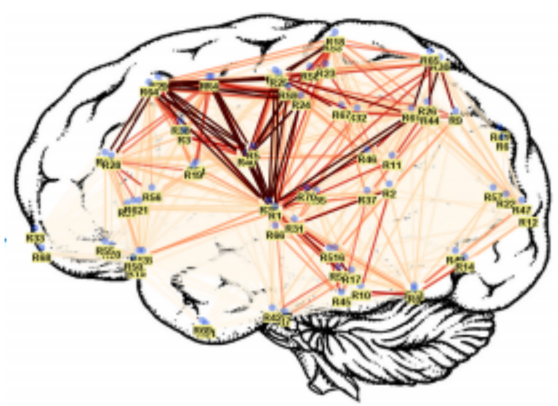
Collaboration Networks



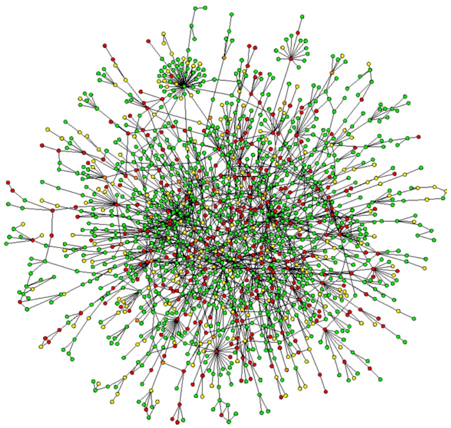
US Power Grid



Traffic Network



Brain Networks



Biological Networks

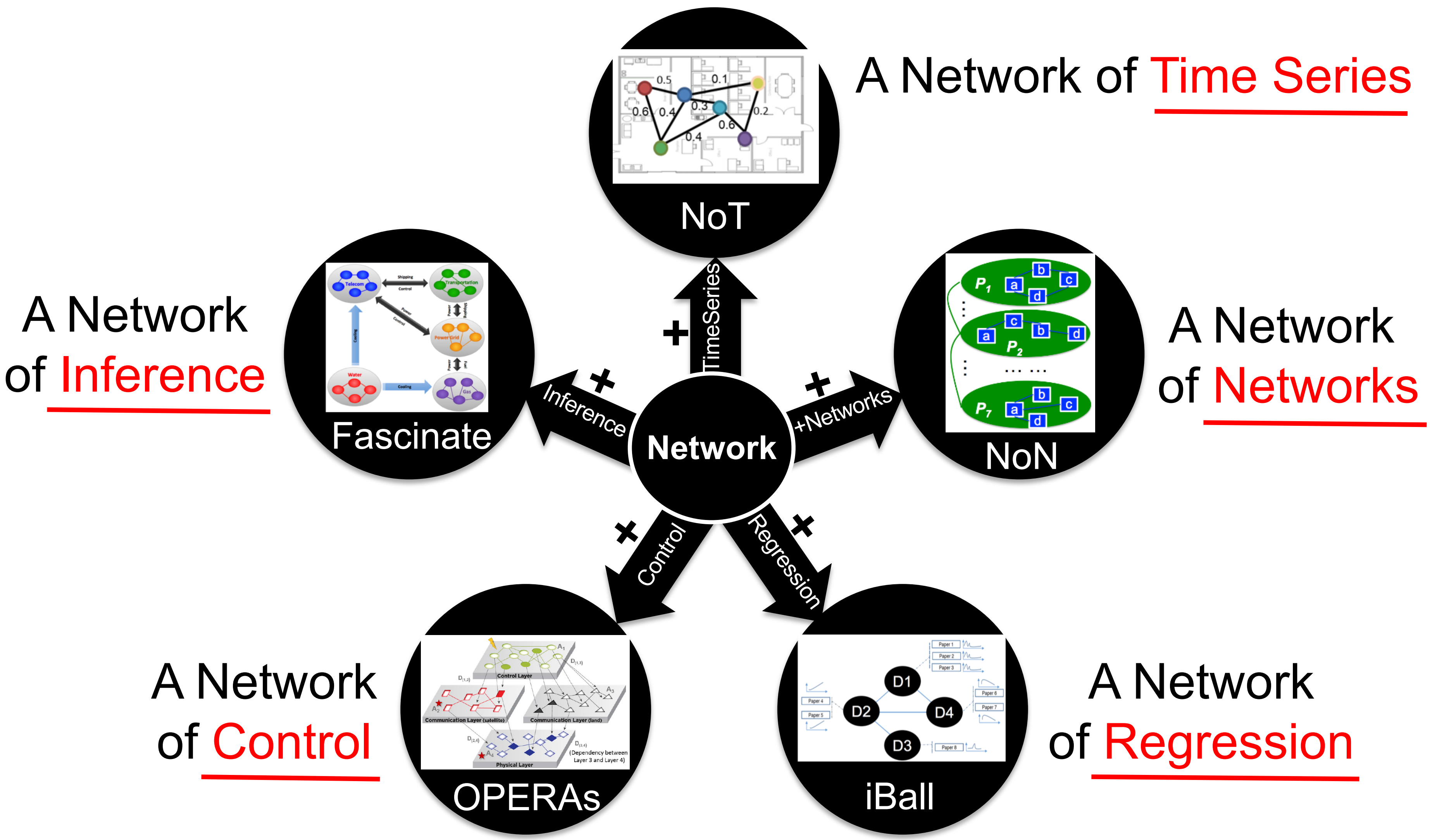
Hospital Networks

This Lecture: Networks = Graphs

Networks-as-Context: Network of X

- **Observation**: Each *X* being an **Entity**
  - X = user: Social Network
  - X = information: Information Network
  - X = webpage: Web Graph
  - X = device: Internet of Things
- **Question**: What if Each *X* is
  - A dataset (a set of entities, e.g., a network), or
  - A data mining model?

Network of X in This Lecture



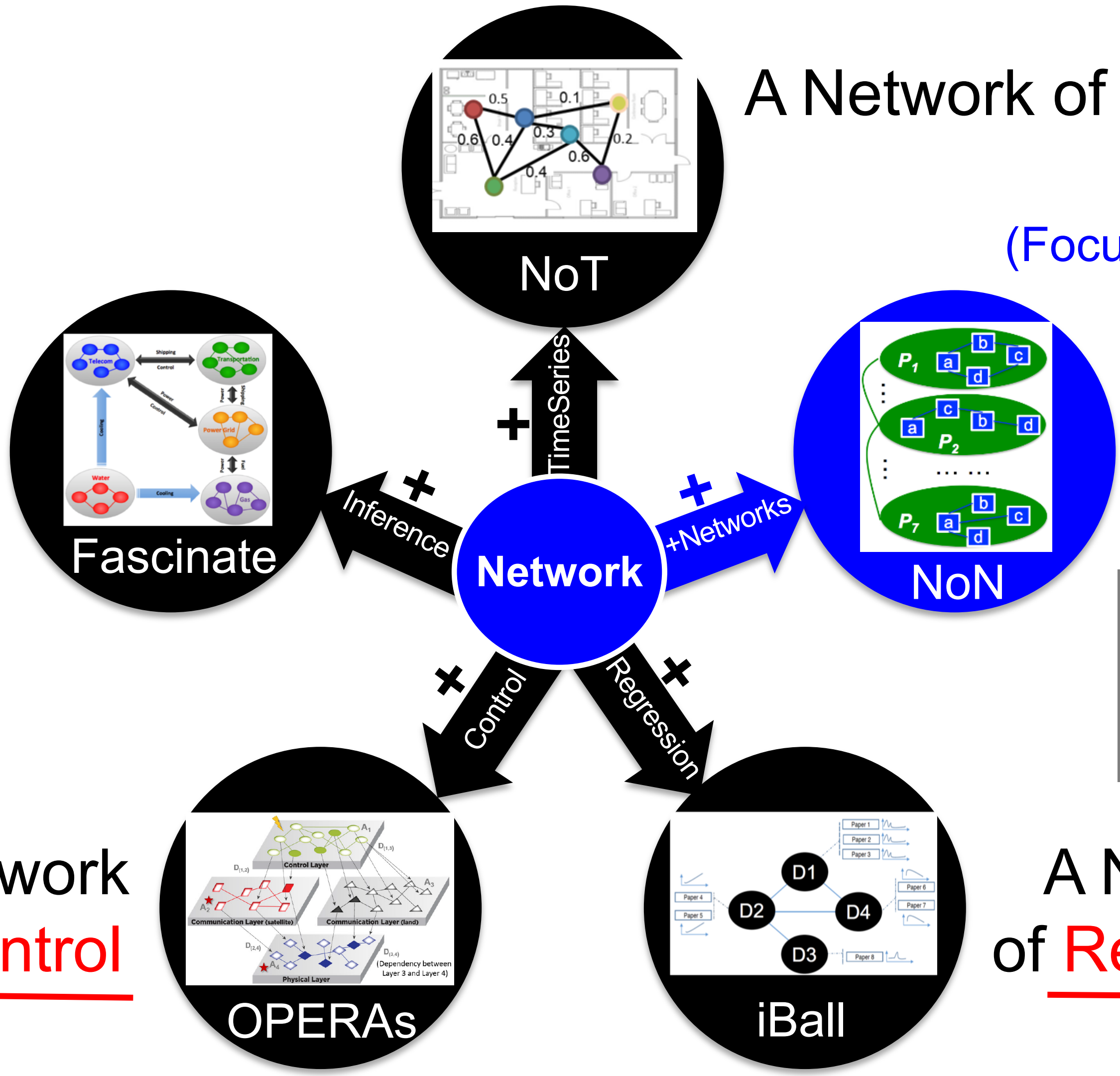
Network of X in This Lecture

A Network of Time Series

(Focus of This Lecture)

A Network of Inference

A Network of Networks



Key Elements in NoN
(a Network of Networks):
(1) Set of networks
(2) Context network
(3) How to map (1) and (2)

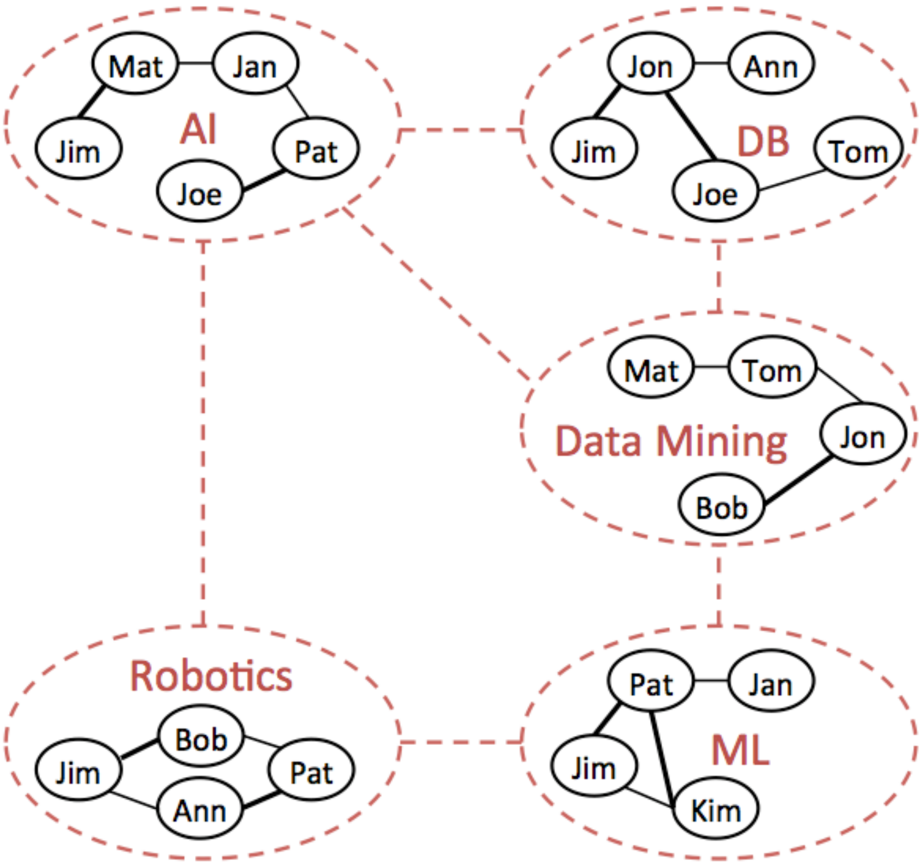
A Network of Control

A Network of Regression

Roadmap

- Motivation
- Network of Networks
  - NoN Introduction
  - NoN Model
  - NoN Construction
  - NoN Mining
- Network of X: Other Scenarios
- Summary

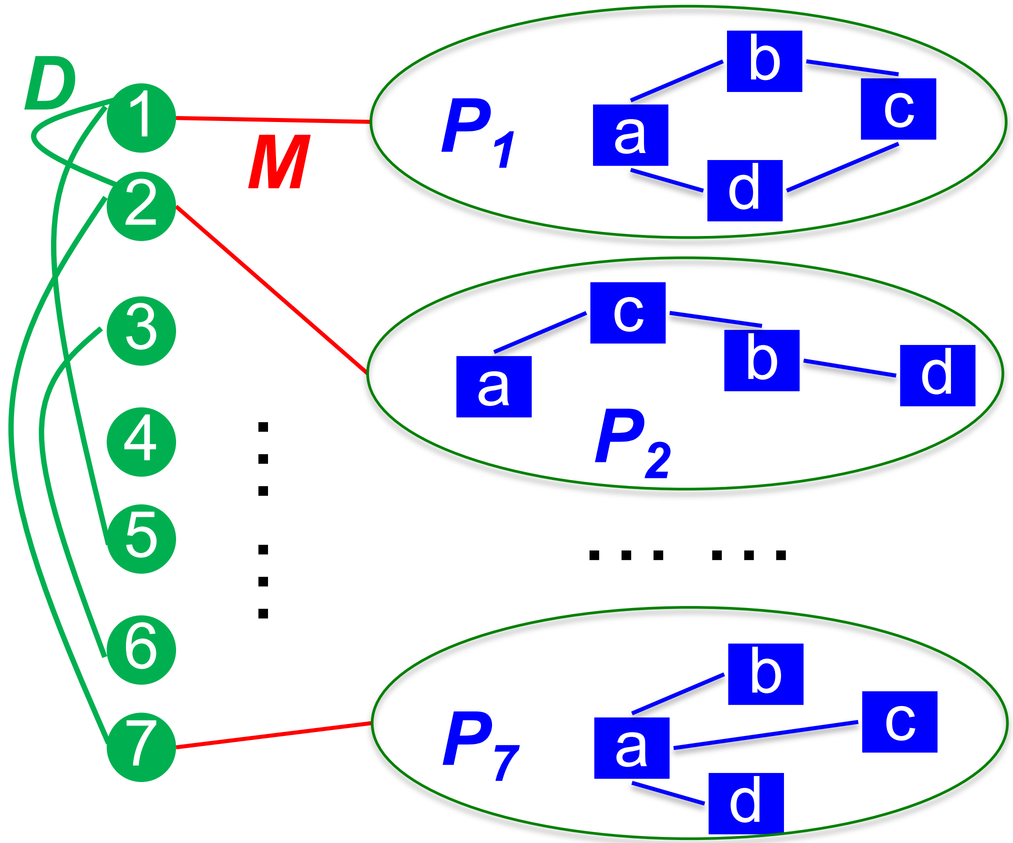
NoN: Scholarly Data Mining



- Set of **co-authorship networks**, connected by **a research domain network**.

- Santo Fortunato et al: Science of science. Science. Vol 359, Issue 6379, 2018
- B. Börner, N. Contractor, et al: :A multi-level systems perspective for the science of team science. Science Translational Medicine, 2(49), 2010
- Y. Lin, H. Tong, J. Tang, K.S. Candan: Guest Editorial: Big Scholar Data Discovery and Collaboration. IEEE Trans. Big Data 3(1): 2 (2017)

Bioinformatics

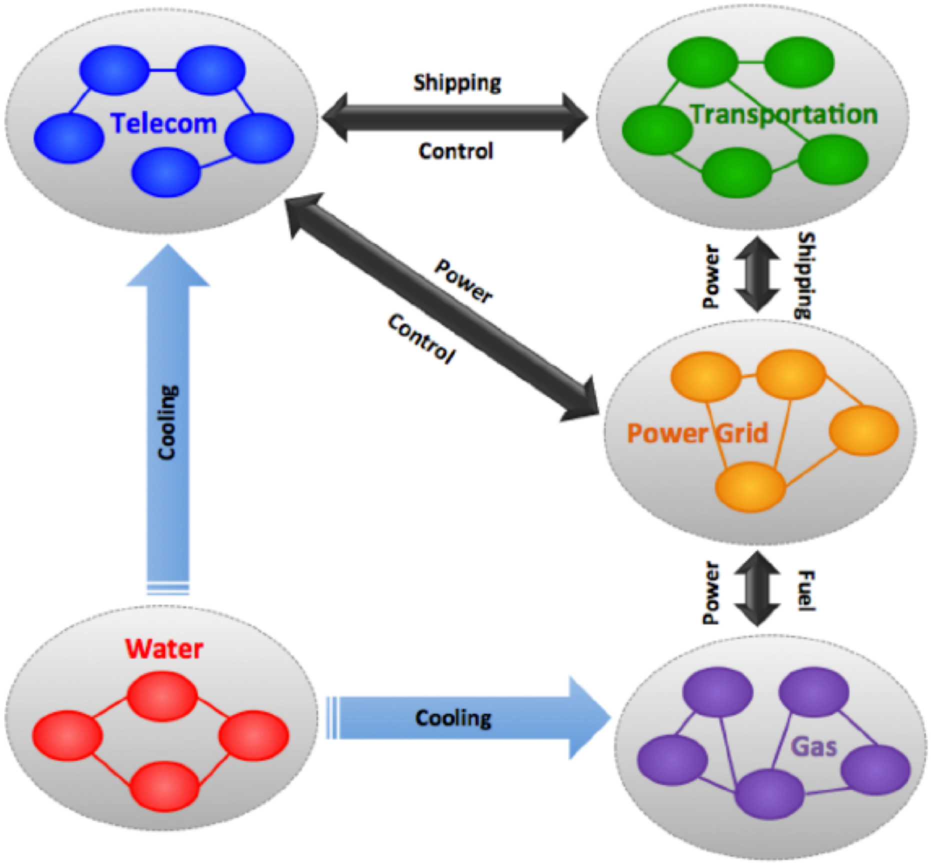


- **Blue**: Tissue-specific PPI networks
- **Green**: Disease similarity network

- Set of **tissue-specific protein-protein interaction (PPI) networks**, connected by **a disease similarity network**.

- O. Magger, Y. Y. Waldman, E. Ruppin, and R. Sharan. Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks. *PLoS Computational Biology*, 8(9), 2012.
- Gross AM, Ideker T. Molecular networks in context. Nature biotechnology. 33(7), 2015

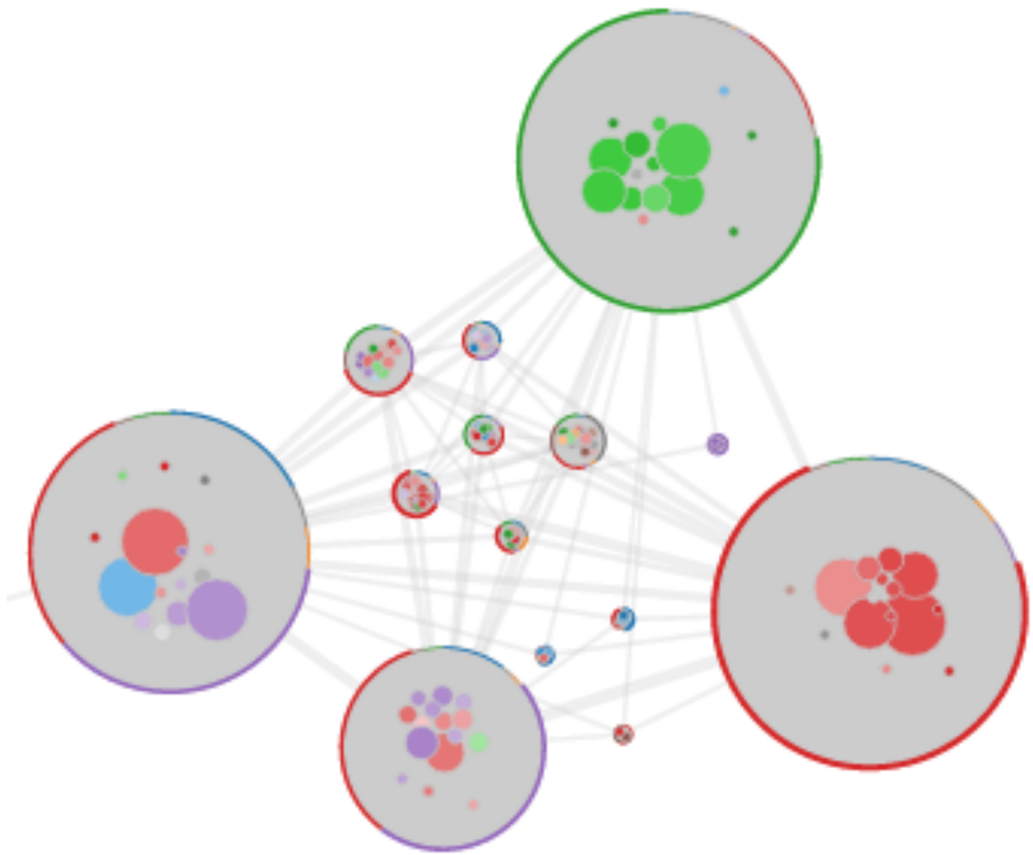
Critical Infrastructure Networks



- Set of **infrastructure networks**, connected by **a cross-infrastructure dependency network**.

- J. Gao, S.V. Buldyrev, H.E. Stanley, S. Havlin: Networks formed from interdependent networks. 8(1) 2012.
- S.D. Reis, Y. Hu, A. Babino, J.S. Andrade, S. Canals, M. Sigman, and H.A. Makse: Avoiding catastrophic failure in correlated networks of networks. Nature Physics, 10(10), 762, 2014
- NSF Critical Resilient Interdependent Infrastructure Systems and Processes (CRISP 2.0): https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505277

Network Science of Teams

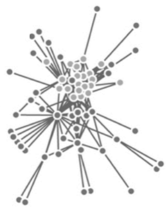
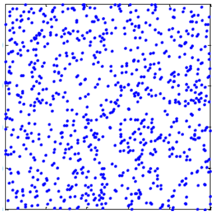
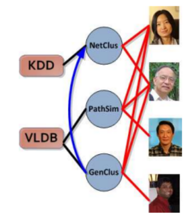


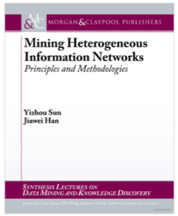
- Set of **team networks**, connected by **a project dependency network**.

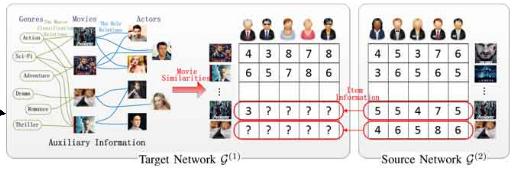
- G.S. McChrystal, C. Tantum, S. David, and F. Chris:. Team of teams: New rules of engagement for a complex world. Penguin, 2015.
- N. Contractor, L.A. DeChurch, A. Sawant, and X. Li: My Dream Team Assembler, 2013.
- W. Stefan, B. Jones, and B. Uzzi: The Increasing Dominance of Teams in the Production of Knowledge. Science, May 2007, 316:1036-1039.
- Network Science of Teams Project Website: http://team-net-work.org
- L. Li, and H. Tong: Network Science of Teams: Characterization, Prediction, and Optimization. WSDM 2018 tutorial

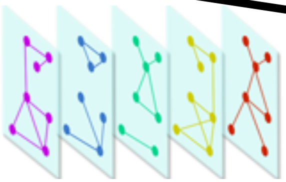
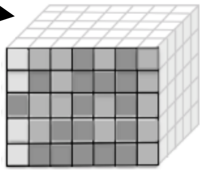
Network of Networks vs. Other Complex Networks

- **Existing Work**
  - Plain Networks
  - Heterogeneous Information Networks [Sun & Han 2012-2017]
  - Multi-Networks [Yu+ 2013-2018]
  - Tensor [Faloutsos+ 2008-2017]
  - Multiplex [Kanawati 2015, Porter 2014]

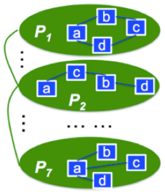
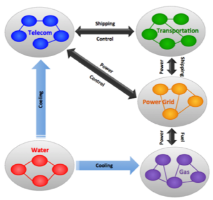
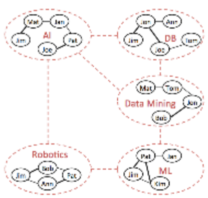








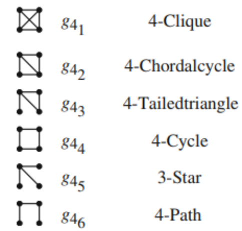



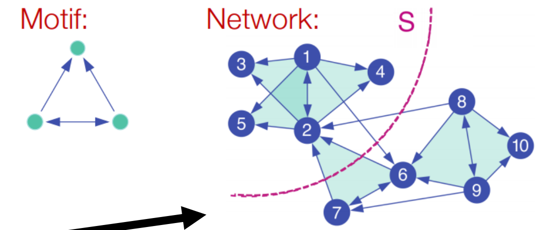
- **NoN**: Set of Networks Connected by a Contextual Network
- **Key Advantages: Network Modeling**
  - Multi-Resolution, Hierarchical Modeling
  - Towards Multi-Network Model Unification
    - Including tensor, multiplex, hypergraph models as special cases

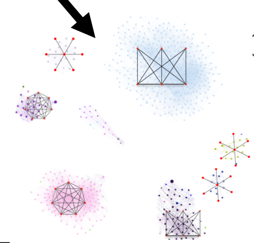
Network of Networks vs. High-order Structure

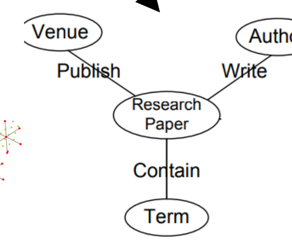
- **Existing Work**
  - A Node or a Link 
  - Graphlet [Neville+ 2016]
  - Motif [Benson+ 2016]
  - Meta-Path or Meta-Structure [Sun & Han 2012-2017]
  - Graph Vocabulary [Koutra & Faloutsos+ 2014]





- **NoN**: Networks Inside a Node/Link
- **Key Advantages: Structure Aggregation**
  - Collective Mining with Focused Knowledge Transfer
  - New Principle for Mining Multiple Networks: Cross Network Consistency

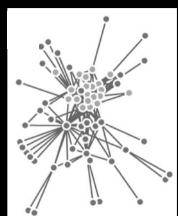
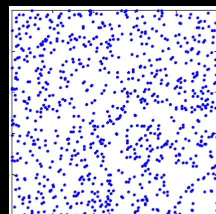




Network of Networks: Where Are We?

Basic Networks

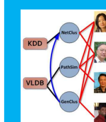
- **At Macro-level:** Nodes Linked by Edges (e.g., an Adjacency Matrix)

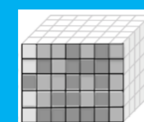
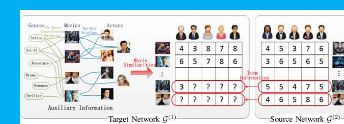



- **At Micro-level:** Node/Link as Atom

Beyond a Single Adj. Matrix

1. HIN [Sun & Han 2012-2017]
2. Tensor [Faloutsos+ 2008-2017]
3. Multi-Network [Yu+ 2013-2018]
4. Multiplex [Kanawati 2015, Porter 2014]



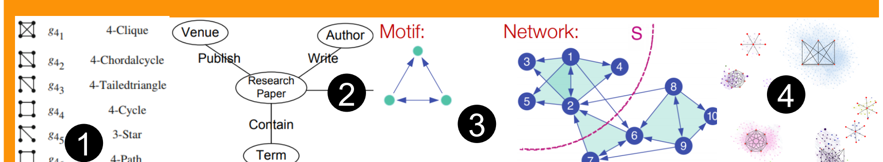
1

2


3

4

High-order Structure

1. Graphlet [Neville+ 2016]
2. Meta-Path/Structure [Sun & Han 2012]
3. Motif/High-order Structure [Benson+ 2016]
4. Graph Vocabulary [Koutra, Faloutsos+ 2014]



- **At Macro-level:** Set of Networks Connected by Another Network
- **Key Advantages:** Hierarchical Modeling, Towards Multi-Network Model Unification.

NoN: Network of Networks

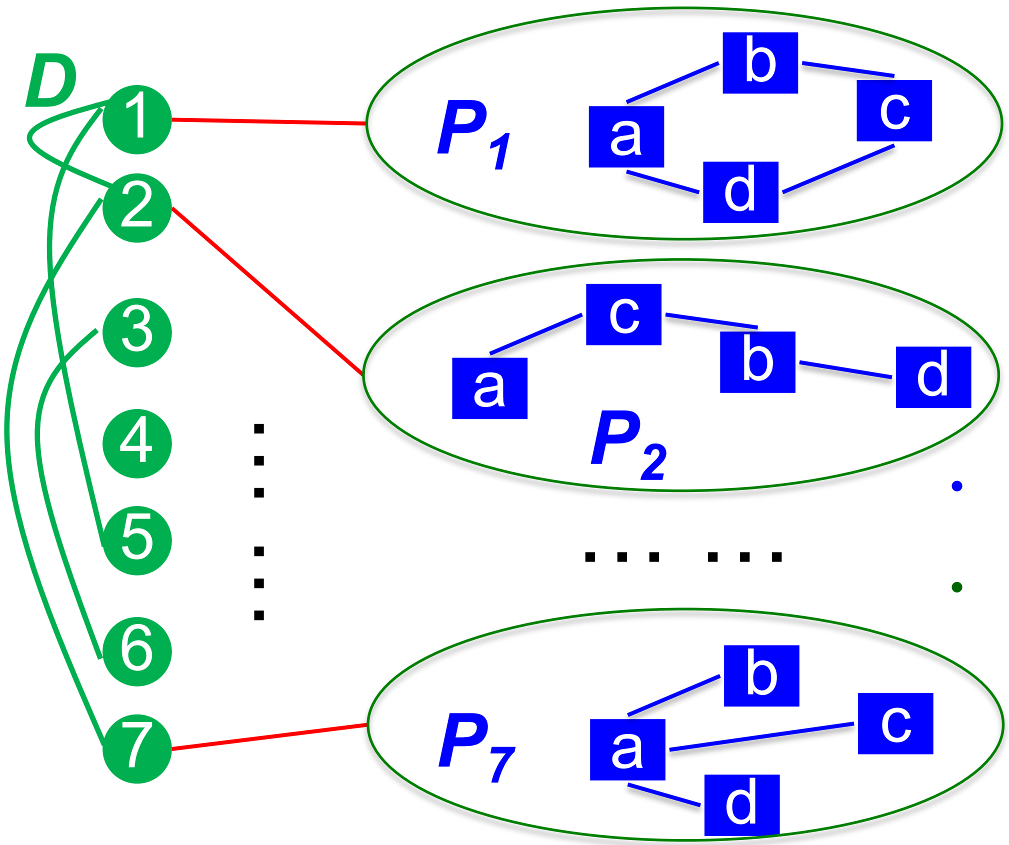
- **At Micro-level:** Hidden Networks Deep Inside a Node/Link
- **Key Advantages**: Collective Mining with Focused Knowledge Transfer

Roadmap

- Motivation
- Network of Networks
  - NoN Introduction
  - NoN Model
  - NoN Construction
  - NoN Mining
- Network of X: Other Scenarios
- Summary

Modeling NoN

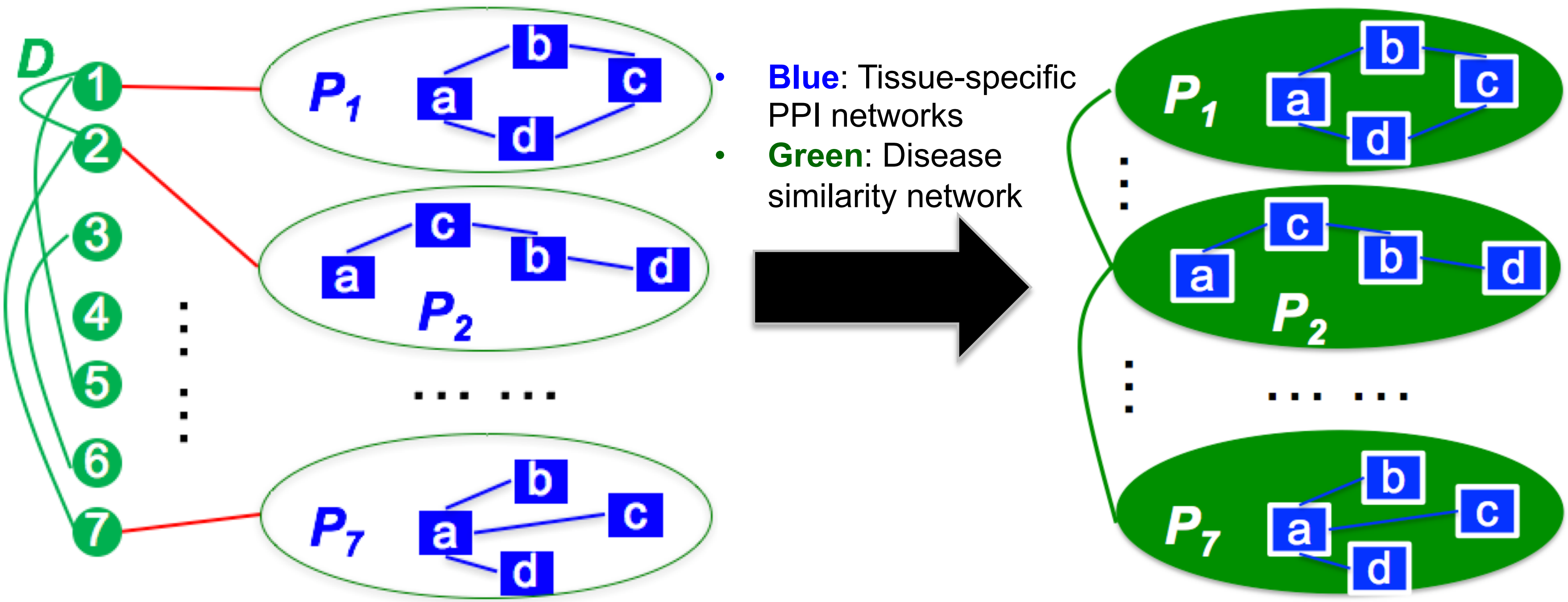
- Q: How to Represent a Set of Inter-connected Networks (e.g., Tissue-Specific PPI Networks)?



- **Blue**: Tissue-specific PPI networks
- **Green**: Disease similarity network

Introducing the NoN Model

- A: Each Green Node (Disease) Itself is a Network.



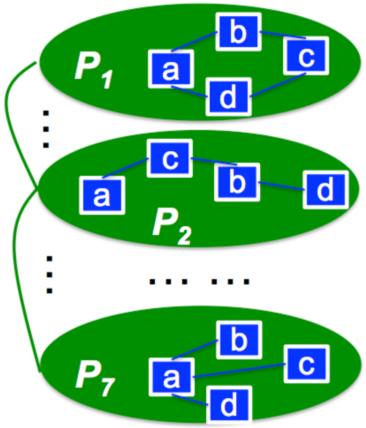
NoN (A Network of Networks) := a triplet $R = <G, A, \theta>$

- G: Main network (the green, disease to disease network)
- A: Domain networks (the blue, tissue-specific PPI networks)
- θ: Mapping function (each green, main node → a blue, domain network)

- J. Ni, H. Tong, W. Fan, X. Zhang: Inside the atoms: ranking on a network of networks. KDD 2014
- J. Ni, M. Koyuturk, H. Tong, J. Haines, R. Xu, X. Zhang: Disease gene prioritization by integrating tissue-specific molecular networks using a robust multi-network model. BMC bioinformatics 2016

NoN Models: Examples

- **Blue**: Tissue-specific PPI networks
- **Green**: Disease similarity network



NoN (A Network of Networks) := a triplet $R = \langle G, A, \theta \rangle$

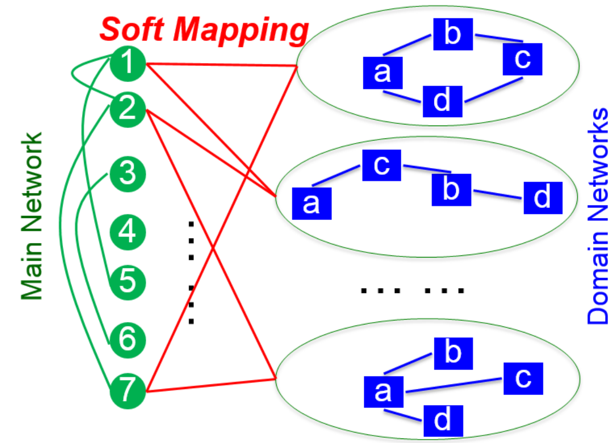
- G: Main network (the green, disease to disease network)
- A: Domain networks (the blue, tissue-specific PPI networks)
- θ: Mapping function (each green, main node → a blue, domain network)

Applications	The Main Network (G)	Domain Networks (A)
Gene-Pheno Assoc.	Disease Similarity Network	Tissue-Specific PPI Nets
Location-Based Social Networks	Geo-Proximity Network	Social Networks
Brain Initiative	Person-Person Network	Brain Networks
Team of Teams	Project Dependence Net	Team Networks
Scholarly Data	Res. Area Similarity Network	Collaboration Networks
Infrastructure Nets	Net-Net Dependence	Gas/Water/Tele/Power Nets

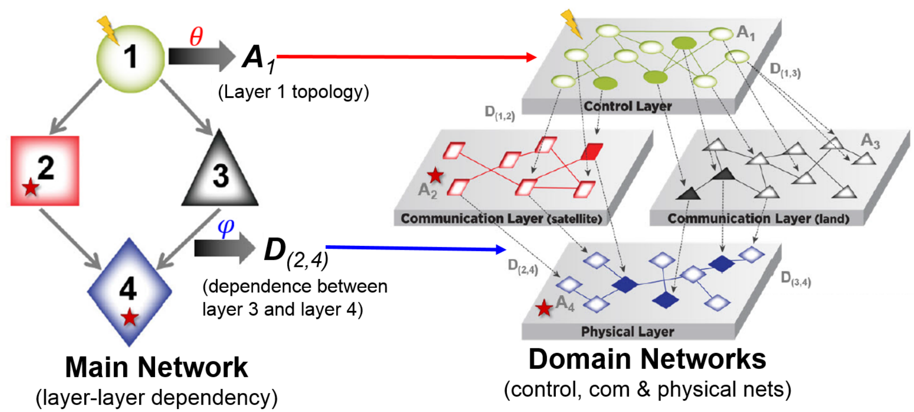
NoN Generalizations

- **G1: Multi-layered Hierarchical NoN**

- **G2: Soft Mapping Function θ**
  - *1-to-many or many-to-many*



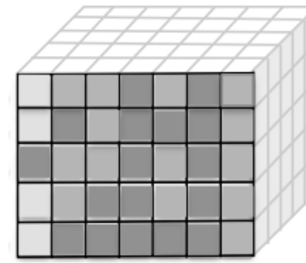
- **G3: Map Edges to Networks φ**
  - G: Main Network
  - $\boldsymbol{A}$: Domain Networks
  - $\boldsymbol{D}$: Cross-Layer Dep'
  - θ: Function $V_G \rightarrow \boldsymbol{A}$
  - φ: Function $E_G \rightarrow \boldsymbol{D}$



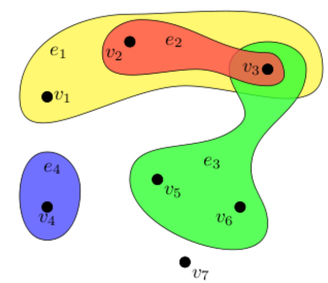
- C. Chen, J. He, N. Bliss and H. Tong: "On the Connectivity of Multi-layered Networks: Models, Measures and Optimal Control" ICDM 2015.
- C. Chen, J. He, N. Bliss and H. Tong: "Towards Optimal Connectivity on Multi-layered Networks". IEEE Trans. Knowl. Data Eng., 29(10): 2332-2346 (2017)

Towards Multi-Network Model Unification

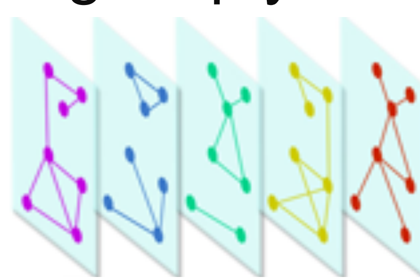
- Special Cases of NoN Model
  - Tensor = NoN with
    1) A full clique main network (*G*)
    2) All domain networks (*A*) sharing the same node sets



  - Hypergraph = NoN with
    1) All domain networks (*A*) being empty



  - Multiplex = NoN with
    1) Two-layers
    2) All domain networks (*A*) sharing the same node sets

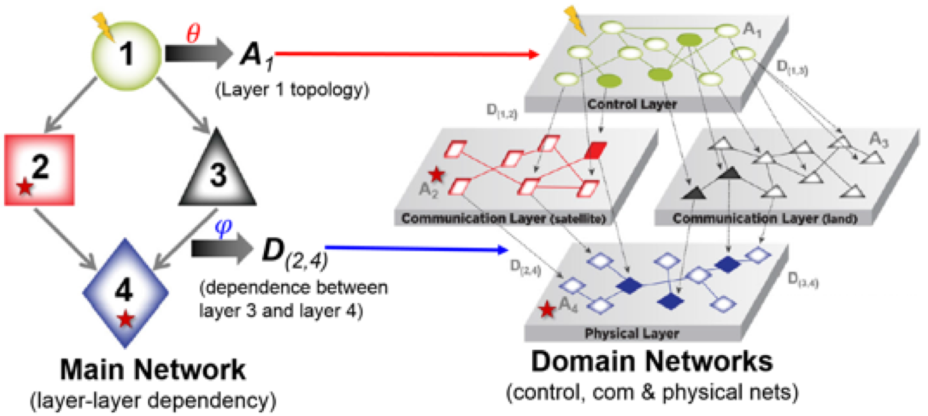


Roadmap

- Motivation
- Network of Networks
  - NoN Introduction
  - NoN Model
  - NoN Construction
  - NoN Mining
- Network of X: Other Scenarios
- Summary

NoN Model Construction

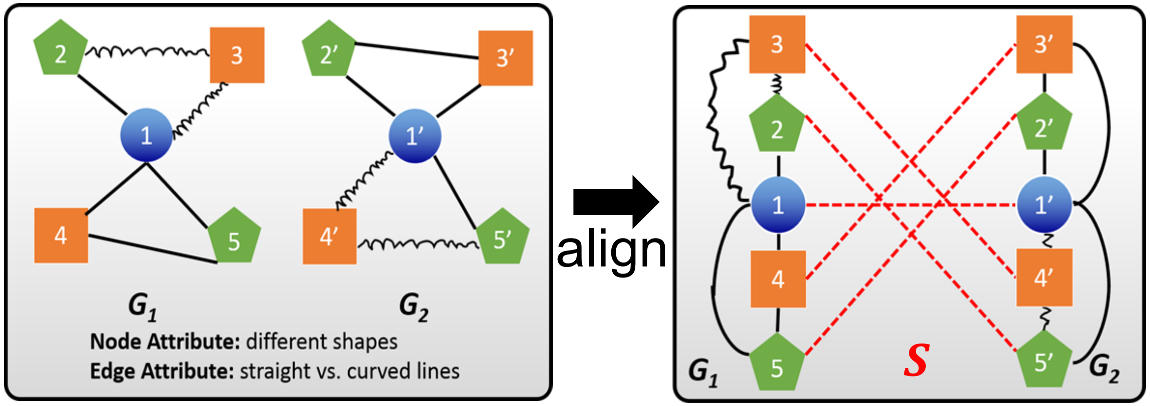
- (Generalized) NoN Model
  - *G*: Main Network
  - ***A***: Domain Networks
  - ***D***: Cross-Layer Dep'
  - θ: Function $V_G \rightarrow$ ***A***
  - φ: Function $E_G \rightarrow$ ***D***



- Q: How to Construct it, Solely Based on *A*?
  - θ function (e.g., main network ***G***)
  - φ function (e.g., cross-layer assn' or dep' ***D***)
- Solution: Soft Network Alignment!
  - **(see lecture 8: network alignment)**

NoN Construction via Soft Network Alignment

- **Soft (Attributed) Network Alignment**
  - **Given**: (1) Two main networks $\{\boldsymbol{A}_1, \boldsymbol{N}_1, \boldsymbol{E}_1\}$ and $\{\boldsymbol{A}_2, \boldsymbol{N}_2, \boldsymbol{E}_2\}$,
    (2) Prior alignment preference $\boldsymbol{H}$ (optional);
  - **Find**: Soft node alignment matrix $\boldsymbol{S}$.
- **Alignment → NoN Model**
  - *S*: Cross-layer association or dependence (φ function)
  - Agg(*S*): Edge in main network
- Illustrative Example



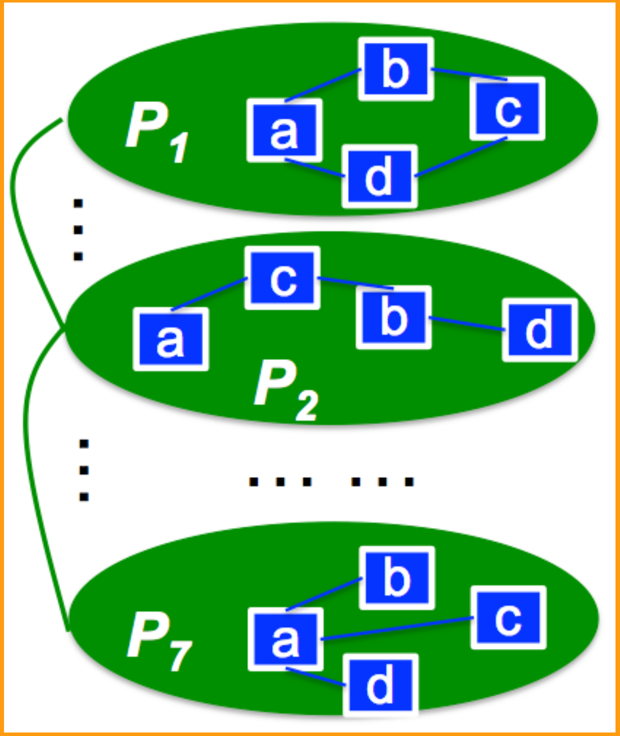
- S. Zhang, and H. Tong: FINAL: Fast Attributed Network Alignment. KDD 2016: 1345-1354
- S. Zhang, H. Tong. Attributed Network Alignment: Problem Definitions and Fast Solutions. IEEE TKDE 31(9): 1680-1692 (2019)

Roadmap

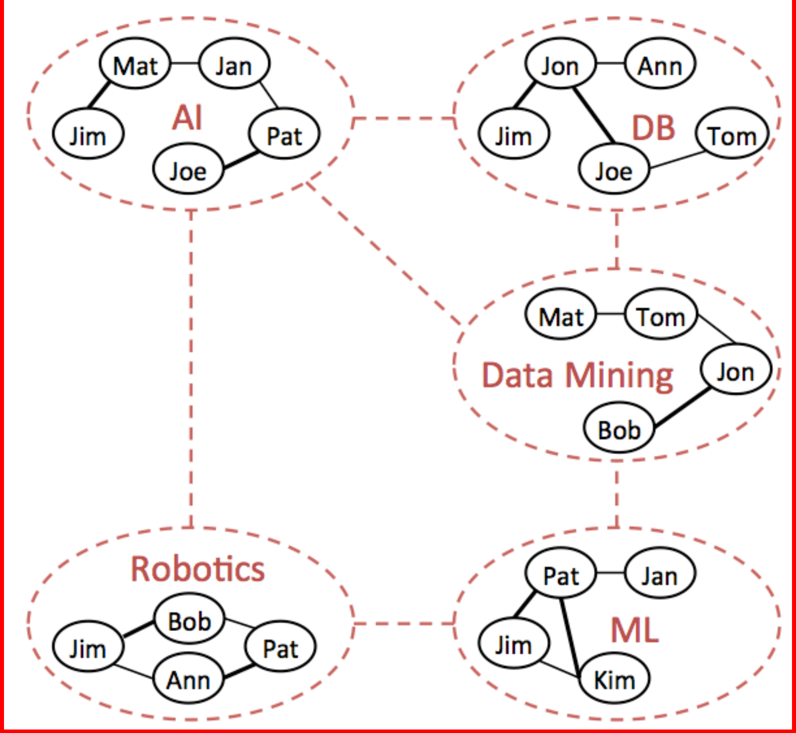
- Motivation
- Network of Networks
  - NoN Introduction
  - NoN Model
  - NoN Construction
  - NoN Mining
- Network of X: Other Scenarios
- Summary

NoN Mining: Ranking

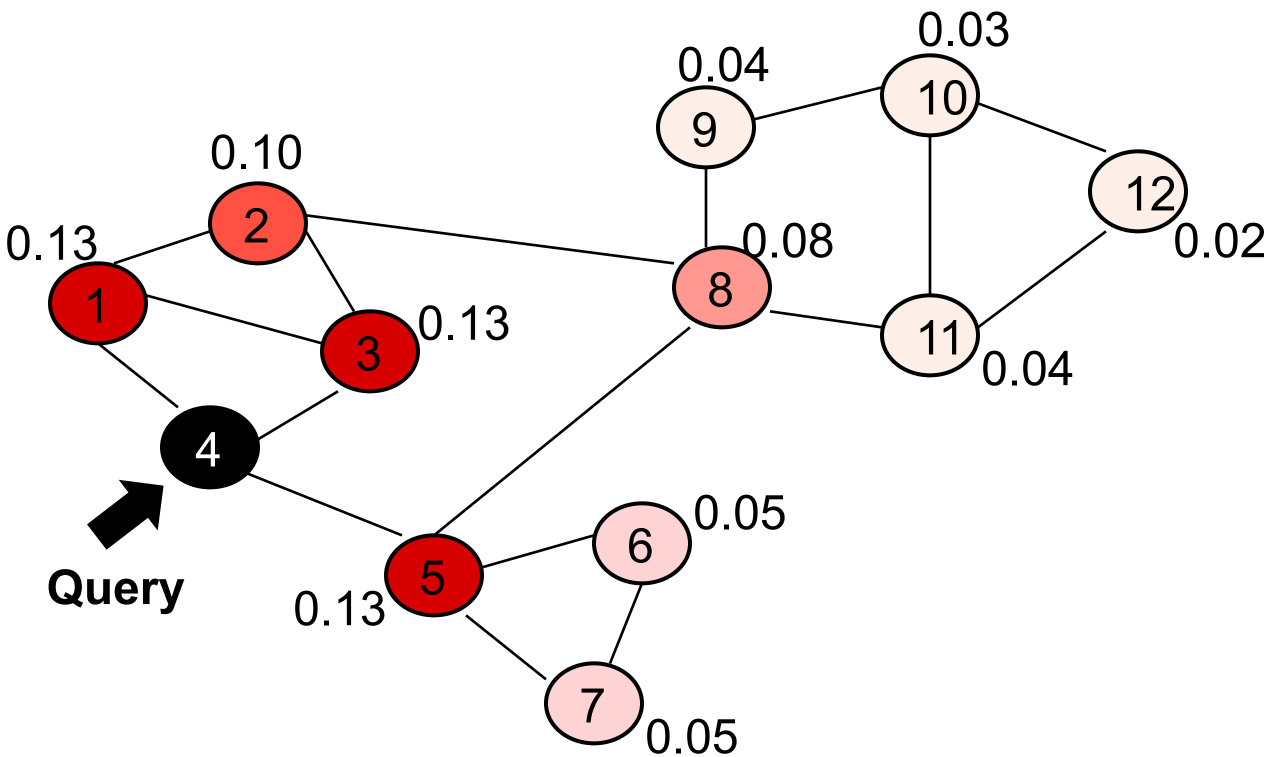
A1: Given a disease (e.g. P_1), what are the most relevant proteins (blue nodes)?



A2: Who is most influential, considering both the within- and cross-area influence?



Ranking on Single Network: Intuition



	Node 4
Node 1	0.13
Node 2	0.10
Node 3	0.13
Node 4	0.22
Node 5	0.13
Node 6	0.05
Node 7	0.05
Node 8	0.08
Node 9	0.04
Node 10	0.03
Node 11	0.04
Node 12	0.02

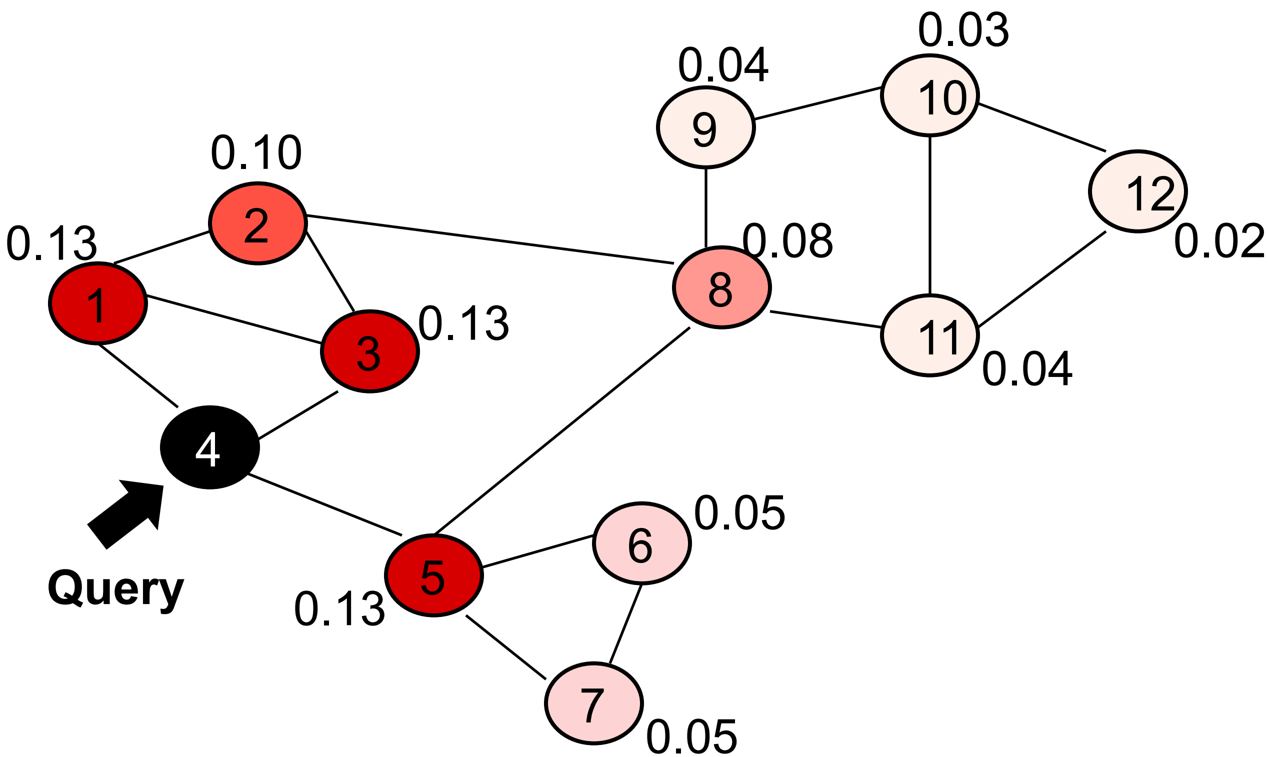
Ranking vector

r_4

- **Assumption**: Homophily (Guilt-by-association)
- **Example**: Two researchers are close if
  - sharing many common co-authors,
  - working on similar topics, and
  - publishing at same venue(s).

H. Tong, C. Faloutsos, J.-Y. Pan: Fast Random Walk with Restart and Its Applications. ICDM 2006
[best paper ICDM 2006, 10-Year Highest Impact Paper Award ICDM 2015]

Ranking on Single Network: Formulation



	Node 4
Node 1	0.13
Node 2	0.10
Node 3	0.13
Node 4	0.22
Node 5	0.13
Node 6	0.05
Node 7	0.05
Node 8	0.08
Node 9	0.04
Node 10	0.03
Node 11	0.04
Node 12	0.02

Nearby nodes, Higher scores

More red, More relevant

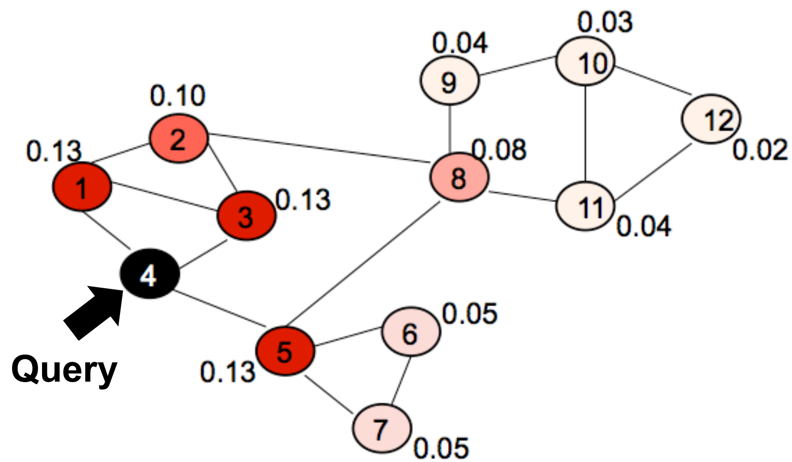
Ranking vector

r_4

$$r_i = c \times A \times r_i + (1-c) \times e_i$$

"Maxwell Equation" for Web -- *Soumen Chakrabarti*

Ranking on Single Network: Optimization



	Node 4
Node 1	0.13
Node 2	0.10
Node 3	0.13
Node 4	0.22
Node 5	0.13
Node 6	0.05
Node 7	0.05
Node 8	0.08
Node 9	0.04
Node 10	0.03
Node 11	0.04
Node 12	0.02

Ranking vector

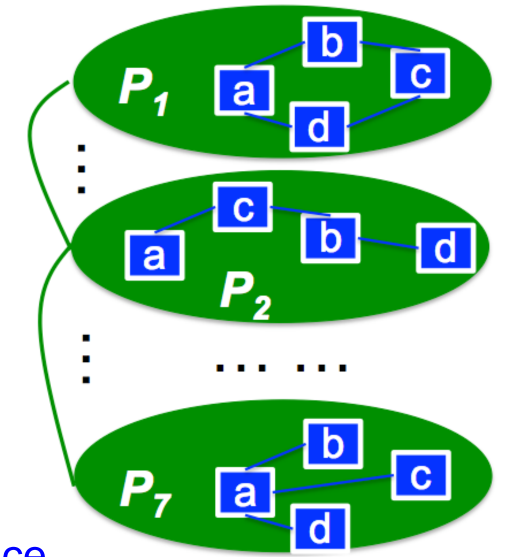
r_4

$$r_i = c \times A \times r_i + (1\text{-}c) \times e_i$$

$$= \mathbf{argmin}\ \underbrace{c r_i'(I - A) r_i}_{\text{Network Smoothness}} + (1\text{-}c) \times \underbrace{\|r_i - e_i\|^2}_{\text{Query Preference}}$$

An Optimization Viewpoint of “Maxwell Equation” for Web (Symmetric *A*)

NoN Ranking: Formulation



- Optimization Formulation:

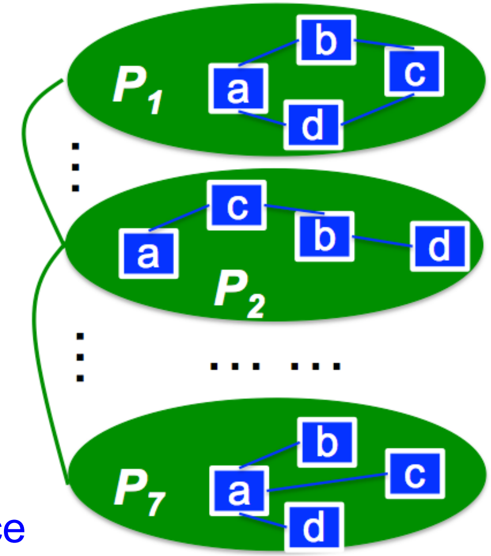
$$
\begin{aligned}
J(\mathbf{r}_1, \ldots, \mathbf{r}_g) \;=\;& c \underbrace{\sum_{i=1}^{g} \mathbf{r}_i'(\mathbf{I}_{n_i} - \tilde{\mathbf{A}}_i)\mathbf{r}_i}_{\text{#1: within-network smoothness}} + (1-c) \underbrace{\sum_{i=1}^{g} \|\mathbf{r}_i - \mathbf{e}_i\|_F^2}_{\text{#2: query preference}} \\
&+ a \underbrace{\sum_{i=1}^{g} \sum_{j=1}^{g} \left\| \frac{\mathbf{r}_i(\mathcal{I}_{ij})}{\sqrt{d_m(i)}} - \frac{\mathbf{r}_j(\mathcal{I}_{ij})}{\sqrt{d_m(j)}} \right\|_F^2 \mathbf{G}(i,j)}_{\text{#3: cross-network consistency}}
\end{aligned}
$$

- *G(i,j)*: similarity between disease *i* and disease *j*
- I_{ij}: overlapped proteins between disease *i* and disease *j*

- Intuition:
  - Similar ranking scores for an overlapped domain node, if their *G(i,j)* is high.
    - **Example**: If (1) a protein is highly relevant to disease-*i*, (2) disease-*i* is very similar to disease-*j*; then it is likely that the same protein is also highly relevant to disease-*j*.
  - A set of correlated *g* random walks

- J. Ni, H. Tong, W. Fan, X. Zhang: Inside the atoms: ranking on a network of networks. KDD 2014
- J. Ni, M. Koyuturk, H. Tong, J. Haines, R. Xu, X. Zhang: Disease gene prioritization by integrating tissue-specific molecular networks using a robust multi-network model. BMC bioinformatics 2016

NoN Ranking: Equivalence



- Optimization Formulation:

$$J(\mathbf{r}_1, \ldots, \mathbf{r}_g) = c \underbrace{\sum_{i=1}^{g} \mathbf{r}_i'(\mathbf{I}_{n_i} - \tilde{\mathbf{A}}_i)\mathbf{r}_i}_{\text{\#1: within-network smoothness}} + (1-c)\underbrace{\sum_{i=1}^{g} \|\mathbf{r}_i - \mathbf{e}_i\|_F^2}_{\text{\#2: query preference}}$$

$$+ a \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{g} \left\| \frac{\mathbf{r}_i(\mathcal{I}_{ij})}{\sqrt{d_m(i)}} - \frac{\mathbf{r}_j(\mathcal{I}_{ij})}{\sqrt{d_m(j)}} \right\|_F^2 \mathbf{G}(i,j)}_{\text{\#3: cross-network consistency}}$$

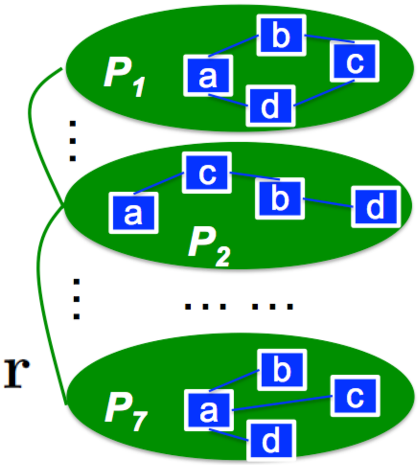
- Equivalence: $J(\mathrm{r}) = J(\mathrm{r}_1, \ldots, \mathrm{r}_g)$

$$J(\mathbf{r}) = c\mathbf{r}'(\mathbf{I}_n - \tilde{\mathbf{A}})\mathbf{r} + (1-c)\|\mathbf{r} - \mathbf{e}\|^2 + 2a\mathbf{r}'(\mathbf{I}_n - \tilde{\mathbf{Y}})\mathbf{r}$$

  – **Intuition**: Single R.W. on the integrated network $\tilde{W}$

- $\tilde{A}$: Diagonal block matrix; Each block is a domain network
- $\tilde{Y}$: Off-diagonal block matrix, Weighted indicator matrix for overlapped node pairs
- $\tilde{W}$: Integrated network, Weighted linear combination of $\tilde{A}$ and $\tilde{Y}$.

NoN Ranking: Algorithms



- **Property**: $J(\mathbf{r})$ is positive-definite!

$$J(\mathbf{r}) = c\mathbf{r}'(\mathbf{I}_n - \tilde{\mathbf{A}})\mathbf{r} + (1-c)\|\mathbf{r} - \mathbf{e}\|^2 + 2a\mathbf{r}'(\mathbf{I}_n - \tilde{\mathbf{Y}})\mathbf{r}$$

- **Proof** (key ideas)
  - Normalized Graph Laplacian: $\mathrm{Eigs}(\tilde{\mathbf{A}}) \in [-1, 1]$ $\mathrm{Eigs}(\tilde{\mathbf{Y}}) \in [-1, 1]$
  - Hessian Matrix: $\nabla^2 J = 2((1+2a)\mathbf{I}_n - c\tilde{\mathbf{A}} - 2a\tilde{\mathbf{Y}})$
  - Weyl's Inequality: $\mathrm{Eigs}(\nabla^2 J) \geq 2(1-c) > 0$
- **Algorithms**:
  - #1: **Linear**, fixed-point algorithm → **optimal** solution

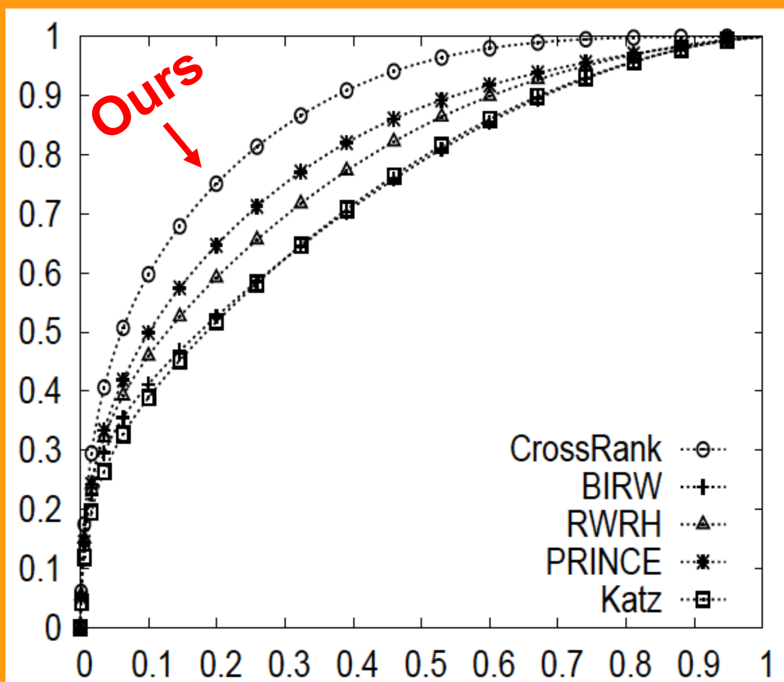
$$\mathbf{r} = (\frac{c}{1+2a}\tilde{\mathbf{A}} + \frac{2a}{1+2a}\tilde{\mathbf{Y}})\mathbf{r} + \frac{1-c}{1+2a}\mathbf{e} \longrightarrow (\mathbf{I}_n - \frac{c}{1+2a}\tilde{\mathbf{A}} - \frac{2a}{1+2a}\tilde{\mathbf{Y}})^{-1}\frac{1-c}{1+2a}\mathbf{e}$$

  - #2: Equivalence → **any** existing fast solution on single network
  - #3: Structure of NoN → further speedup $O(T(m+ng))$ → $O(T(g\log(g) + z))$
    - $g \ll n$; and $z \ll m$ (key idea: using main network for pruning)

NoN Ranking: Results

A1: Candidate Protein Prioritization

- *Which proteins are most relevant w.r.t. disease a?*



ROC Curve Comparison

A2: Collaboration Prediction

- *Which data mining authors are likely to collaborate with a given medical author?*

#Papers	Hops	#Pairs	Methods	AUC	Accuracy
≥ 3	[3, 4]	45	PC	0.7196	0.4444
			Katz	0.7439	0.5556
			PropFlow	0.7558	0.6222
			PathSim	0.5636	0.2444
			PageRank	0.7417	0.5333
			CrossQuery	**0.7685**	**0.6444**
≥ 3	[3, 6]	70	PC	0.6009	0.3000
			Katz	0.6243	0.3714
			PropFlow	0.6268	0.4429
			PathSim	0.5278	0.2143
			PageRank	0.6378	0.3714
			CrossQuery	**0.6632**	**0.4571**
≥ 5	[3, 4]	23	PC	0.6521	0.2609
			Katz	0.6717	**0.3478**
			PropFlow	0.6850	**0.3478**
			PathSim	0.4279	0.1304
			PageRank	0.6743	**0.3478**
			CrossQuery	**0.7099**	**0.3478**

AUC and Accuracy

- Solely based on homogeneous networks
- Tailored for cross-area prediction

NoN Mining: Clustering

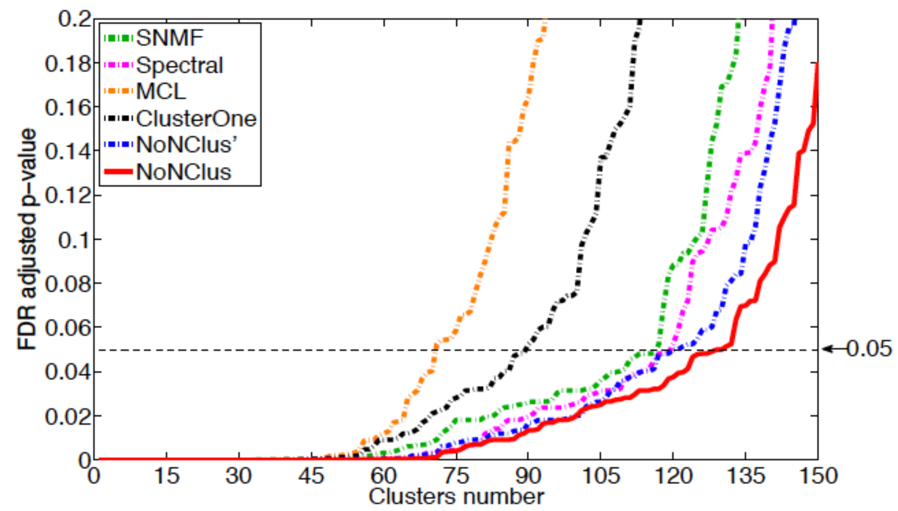
- Obj. Function:

$$\min_{\substack{\mathbf{U}^{(i)} \geq 0 \ (i=1,\ldots,g) \\ \mathbf{V}^{(j)} \geq 0 \ (j=1,\ldots k)}} J_D = \underbrace{\sum_{i=1}^{g} \|\mathbf{A}^{(i)} - \mathbf{U}^{(i)}(\mathbf{U}^{(i)})'\|_F^2}_{domain-specific\ network\ clustering} + a \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{k} h_{ij} \|\mathbf{U}^{(i)} - \mathbf{V}^{(j)}\|_F^2}_{main\ cluster\ guided\ regularization}$$

Same assumption, Similar intuition!

- Algorithm:
  - Linear, multiplicative algorithm converges to KKT point.

- Results:



P-value vs. (biologically meaningful) clusters

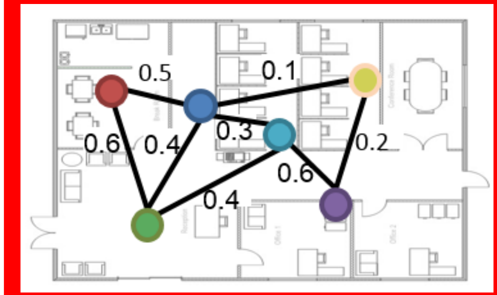
J. Ni, H. Tong, W. Fan, X. Zhang: Flexible and Robust Multi-Network Clustering. KDD 2015

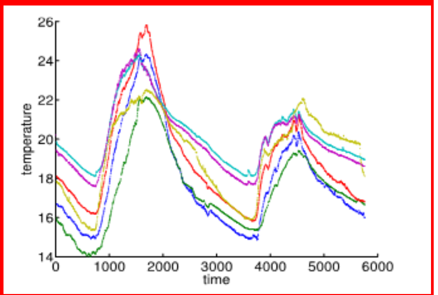
Roadmap

- Motivation
- Network of Networks
- Network of X: Other Scenarios
  - NoT: A Network of Co-evolving Time Series
  - iBall: A Network of Regression Models
  - Fascinate: A Network of Inference Problems
  - Mulan: A Network of Control Problems
- Summary

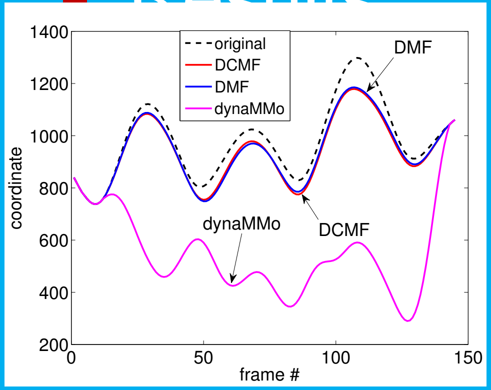
NoT: A Network of Time Series

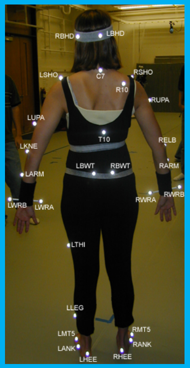
Problem Definition



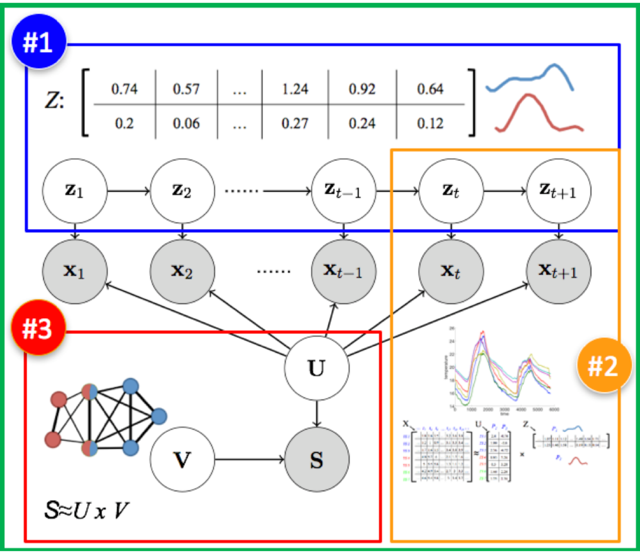


Results





Models & Algorithms



$$\underset{Z,U,V,\theta}{\operatorname{argmax}}\, p(Z,X,V,S,U) = p(z_1)\underbrace{\prod_{t=2}^{T} p(z_t|z_{t-1})}_{\text{temporal smoothness}}$$

#1

$$\times \underbrace{\prod_{t=1}^{T} p(x_t|z_t,U)}_{\text{observed time series}}$$

#2

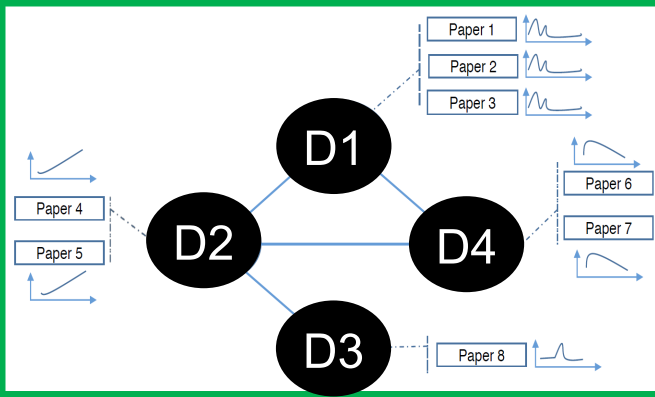
$$\times \prod_{j=1}^{n} p(v_j) \underbrace{\prod_{j=1}^{n} p(s_j|v_j,U)}_{\text{contextual network}}$$

#3

- Y. Cai, H. Tong, W. Fan and P. Ji: Fast Mining of a Network of Coevolving Time Series. *SDM 2015.*
- Y. Cai, H. Tong, W. Fan, P. Ji, Q. He: Facets: Fast Comprehensive Mining of Coevolving High-order Time Series. KDD 2015
- F. Hairi, H. Tong, and L. Ying, NetDyna: Mining Networked Coevolving Time Series with Missing Values. BigData 2019
- Baoyu Jing, Hanghang Tong and Yada Zhu: Network of Tensor Time Series. TheWebConf 2021

iBall: A Network of Regression Models

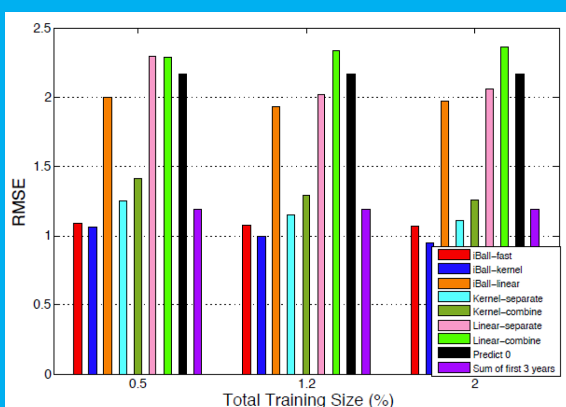
Problem Definition

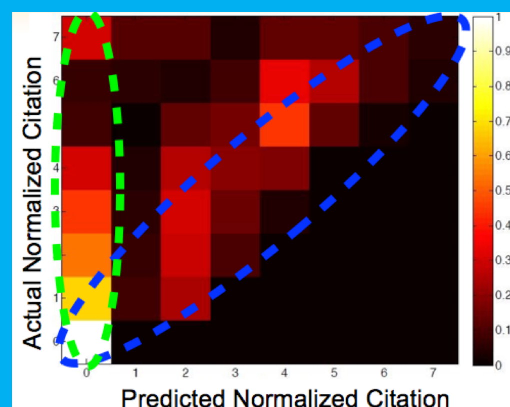


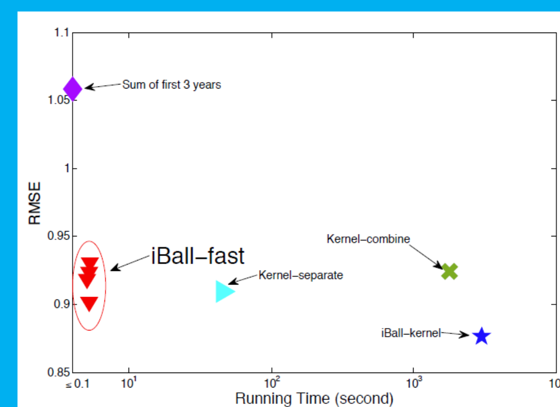
Models & Algorithms

$$
\begin{aligned}
\min_{\mathbf{w}^{(\mathrm{i})},i=1,\ldots,n_d} \quad & \sum_{i=1}^{n_d} \|\mathbf{K}^{(\mathrm{i})}\mathbf{w}^{(\mathrm{i})} - \mathbf{Y}^{(\mathrm{i})}\|_2^2 \\
& + \theta \sum_{i=1}^{n_d}\sum_{j=1}^{n_d} \mathbf{A}_{\mathbf{ij}} \|\mathbf{K}^{(\mathrm{i})}\mathbf{w}^{(\mathrm{i})} - \mathbf{K}^{(\mathrm{ij})}\mathbf{w}^{(\mathrm{j})}\|_2^2 \\
& + \lambda \sum_{i=1}^{n_d} \mathbf{w}^{(\mathrm{i})\prime}\mathbf{K}^{(\mathrm{i})}\mathbf{w}^{(\mathrm{i})}
\end{aligned}
$$

Results





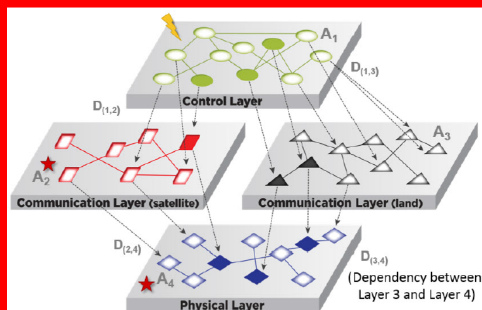


- L. Li, H. Tong and H. Liu: Towards Explainable Networked Prediction. CIKM 2018
- Y. Yao, H. Tong, F. Xu, J. Lu: Scalable Algorithms for CQA Post Voting Prediction. TKDE 2017
- L. Li, H. Tong, Y. Wang, C. Shi, N. Cao and N. Buchler: Is the Whole Greater Than the Sum of Its Parts? KDD 2017
- L. Li, H. Tong: The Child is Father of the Man: Foresee the Success at the Early Stage. KDD *2015*
- Y. Yao, H. Tong, F. Xu, J. Lu: Predicting long-term impact of CQA posts: a comprehensive viewpoint. KDD 2014
- “Data Mining Reveals the Secret to Getting Good Answers”, MIT Technology Review, 2013

OPERAs: Optimal Connectivity Control
(on Multi-Layered Networks)

Problem Definition

Find optimal node set in control layer → w/ maximal impact in target layers



Connectivity Unification

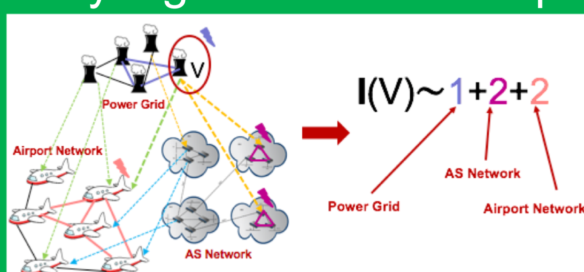
- **Key Idea**: Subgraph aggregation

$$C(\mathbf{A}) = \sum_{\pi \subseteq \mathbf{A}} f(\pi)$$

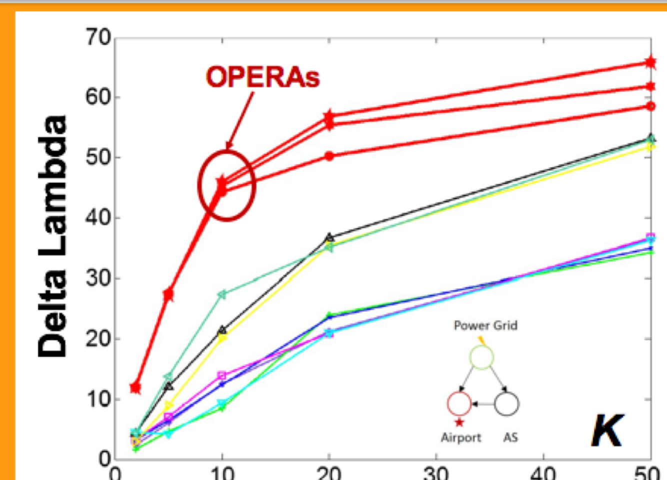
- **Examples:** Path capacity, Robustness, Triangle counts, etc.

Methods

- Impact score: Diminishing returns
- Greedy algorithm → near-optimal



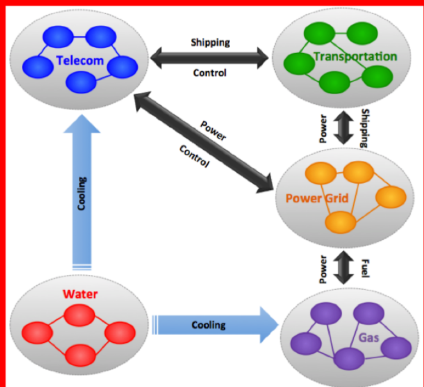
Results

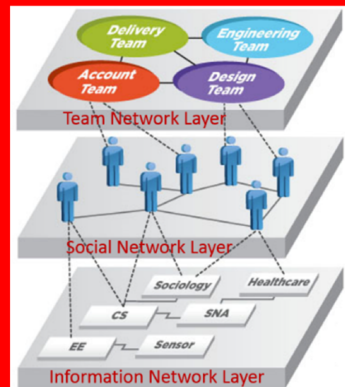


- C. Chen, J. He, N. Bliss and H. Tong: On the Connectivity of Multi-layered Networks: Models, Measures and Optimal Control ICDM15.
- C. Chen, J. He, N. Bliss, H. Tong: Towards Optimal Connectivity on Multi-Layered Networks. IEEE TKDE 2017

Fascinate: Cross-Layer Dependence Inference

Problem Definition





Infer Unobserved Cross-Layer Links

Methods

$$J = \underbrace{\sum_{i,j:\ \mathbf{G}(i,j)=1} \|\mathbf{W}_{i,j} \odot (\mathbf{D}_{i,j} - \mathbf{F}_i{\mathbf{F}_j}')\|_F^2}_{\text{C1: Matching Observed Cross-Layer Dependence}} \quad (1)$$

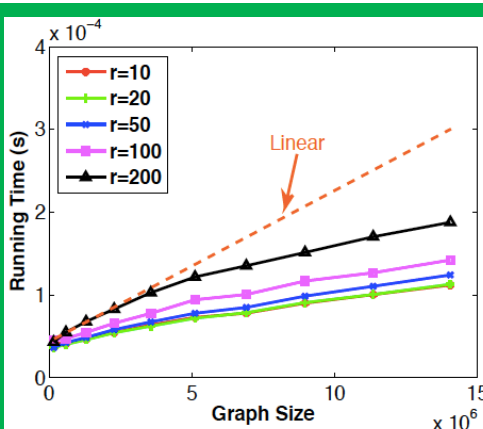
$$+ \underbrace{\alpha \sum_{i=1}^{g} \mathrm{tr}({\mathbf{F}_i}'(\mathbf{T}_i - \mathbf{A}_i)\mathbf{F}_i)}_{\text{C2: Node Homophily}} + \underbrace{\beta \sum_{i=1}^{g} \|\mathbf{F}_i\|_F^2}_{\text{C3: Regularization}}$$

Cross-Layer Inference = Collective CF

Results

Methods	MAP	R-MPR	HLU	AUC	Prec@10
FASCINATE	0.0660	**0.2651**	**8.4556**	**0.7529**	**0.0118**
FASCINATE-CLUST	**0.0667**	0.2462	8.2160	0.7351	0.0108
MulCol	0.0465	0.2450	6.0024	0.7336	0.0087
PairSid	0.0308	0.1729	3.8950	0.6520	0.0062
PairCol	0.0303	0.1586	3.7857	0.6406	0.0056
PairNMF	0.0053	0.0290	0.5541	0.4998	0.0007
PairRec	0.0056	0.0435	0.5775	0.5179	0.0007
FlatNMF	0.0050	0.0125	0.4807	0.5007	0.0007
FlatRec	0.0063	0.1009	0.6276	0.5829	0.0009

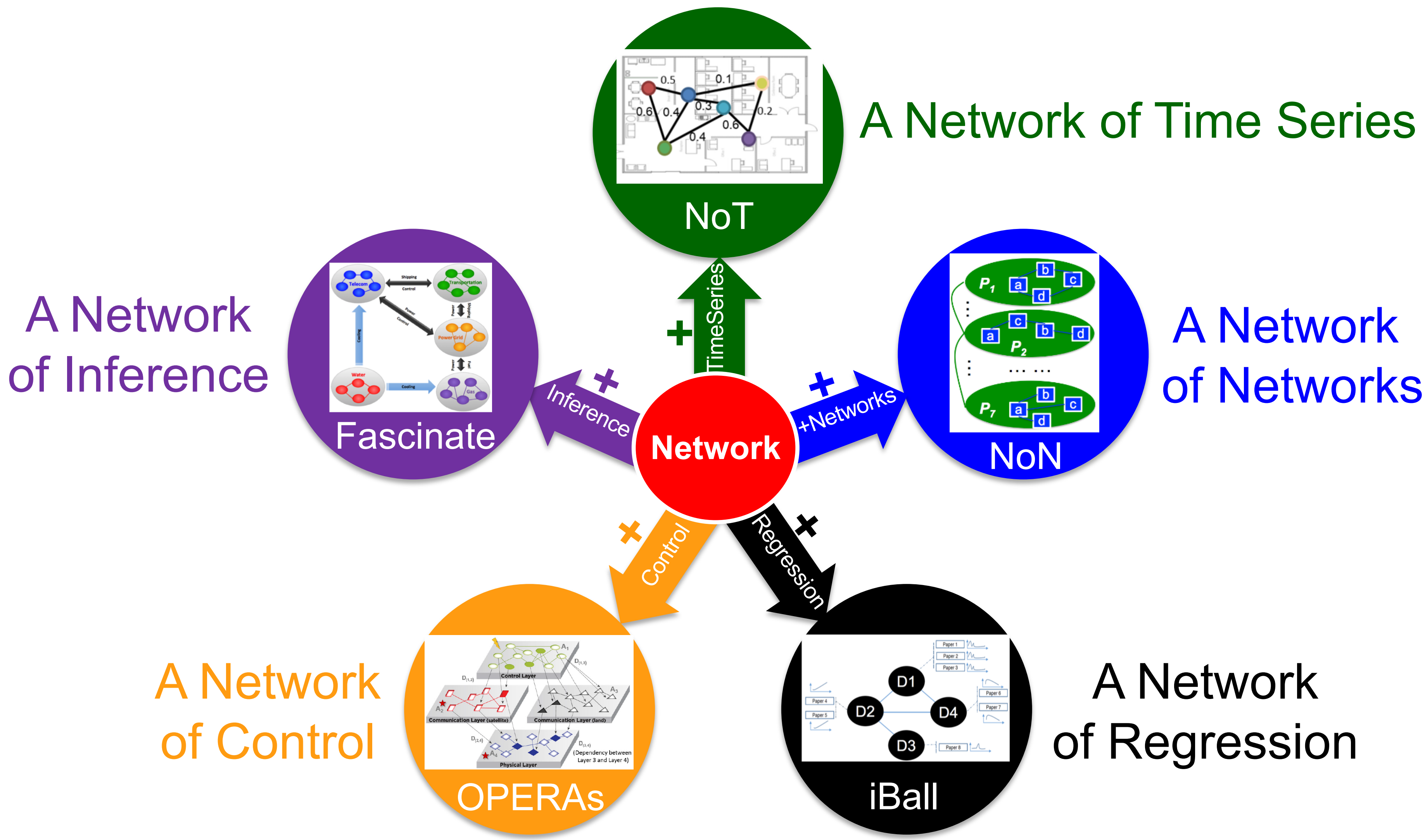
Effectiveness



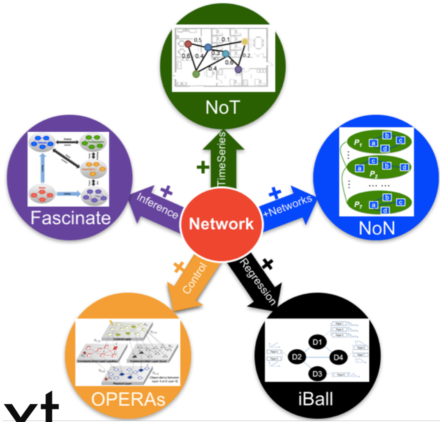
Efficiency

- C. Chen, H. Tong, L. Xie, L. Ying and Q. He: "FASCINATE: Fast Cross-Layer Dependence Inference on Multi-layered Networks", KDD16, 'Bests of KDD 2016'
- C. Chen, H. Tong, L. Xie, L. Ying and Q. He: Cross-Dependency Inference in Multi-layered Networks: A Collaborative Filtering Perspective. ACM TKDD 2017

Summary: Network of X



Summary: Network of X



- **Modeling**: \`**No**' (i.e., Network of X)
  - Networks as data → Networks as context
- **Algorithms**: Networks as Contextual Regularizer
  - e.g., $$J(\mathbf{r}_1, \ldots, \mathbf{r}_g) = \underbrace{c \sum_{i=1}^{g} \mathbf{r}_i'(\mathbf{I}_{n_i} - \tilde{\mathbf{A}}_i)\mathbf{r}_i}_{\text{\#1: within-network smoothness}} + \underbrace{(1-c)\sum_{i=1}^{g} \|\mathbf{r}_i - \mathbf{e}_i\|_F^2}_{\text{\#2: query preference}} + a \underbrace{\sum_{i=1}^{g}\sum_{j=1}^{g} \left\| \frac{\mathbf{r}_i(\mathcal{I}_{ij})}{\sqrt{d_m(i)}} - \frac{\mathbf{r}_j(\mathcal{I}_{ij})}{\sqrt{d_m(j)}} \right\|_F^2 \mathbf{G}(i,j)}_{\text{\#3: cross-network consistency}}$$
- **Computations**: Network for Pruning
  - e.g., O(T(m+ng)) → O(T(g log(g) + z), (g<<n; z<<m)

Reference

- J. Ni, H. Tong, W. Fan, X. Zhang: Inside the atoms: ranking on a network of networks. KDD 2014
- J. Ni, H. Tong, W. Fan, X. Zhang: Flexible and Robust Multi-Network Clustering. KDD 2015
- Y. Cai, H. Tong, W. Fan and P. Ji: Fast Mining of a Network of Coevolving Time Series. *SDM 2015*.
- Y. Cai, H. Tong, W. Fan, P. Ji, Q. He: Facets: Fast Comprehensive Mining of Coevolving High-order Time Series. KDD 2015
- L. Li, H. Tong: The Child is Father of the Man: Foresee the Success at the Early Stage. KDD 2015
- C. Chen, J. He, N. Bliss and H. Tong: On the Connectivity of Multi-layered Networks: Models, Measures and Optimal Control ICDM15.
- L. Li, H. Tong, Y. Wang, C. Shi, N. Cao and N. Buchler: Is the Whole Greater Than the Sum of Its Parts? KDD 2017

Reference

- C. Chen, J. He, N. Bliss, H. Tong: Towards Optimal Connectivity on Multi-Layered Networks. IEEE TKDE 2017
- C. Chen, H. Tong, L. Xie, L. Ying and Q. He: “FASCINATE: Fast Cross-Layer Dependence Inference on Multi-layered Networks”, KDD16, ‘Bests of KDD 2016’
- J. Ni, M. Koyuturk, H. Tong, J. Haines, R. Xu, X. Zhang: Disease gene prioritization by integrating tissue-specific molecular networks using a robust multi-network model. BMC bioinformatics 2016
- C. Chen, H. Tong, L. Xie, L. Ying and Q. He: Cross-Dependency Inference in Multi-layered Networks: A Collaborative Filtering Perspective. ACM TKDD 2017
- F. Hairi, H. Tong, and L. Ying, NetDyna: Mining Networked Coevolving Time Series with Missing Values. BigData 2019
- Baoyu Jing, Hanghang Tong and Yada Zhu: Network of Tensor Time Series. TheWebConf 2021