ONE WORLD SIGNAL PROCESSING SEMINAR SERIES

AUGUST 2020

DELFT UNIVERSITY OF TECHNOLOGY

Acknowledgements:

Mario Coutino (TUDelft)
Georgios B. Giannakis (UofM)
Georgios V. Karanikolas (UofM)
Qiuling Yang (UofM)
Gang Wang (UofM)

GEERT LEUS
DELFT UNIVERSITY OF TECHNOLOGY

LEARNING OF HIGHER-ORDER INTERACTIONS

WITH GRAPH VOLTERRA MODELS

Networks are everywhere

Correspondence

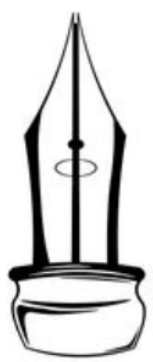
nodes : accounts
edges : communication

Social interactions

nodes : users
edges : interactions

Drug compounds

nodes : substances
edges : same drug



Authorship

nodes : authors
edges : collaboration

Most of them exhibit "higher-order" interactions

Correspondence

nodes : accounts
emails have many recipients

Social interactions

nodes : users
people gather in small groups

Drug compounds

nodes : substances
several substances in a drug

Authorship

nodes : authors
papers have several authors

Today:

How can we learn such higher-order interactions?

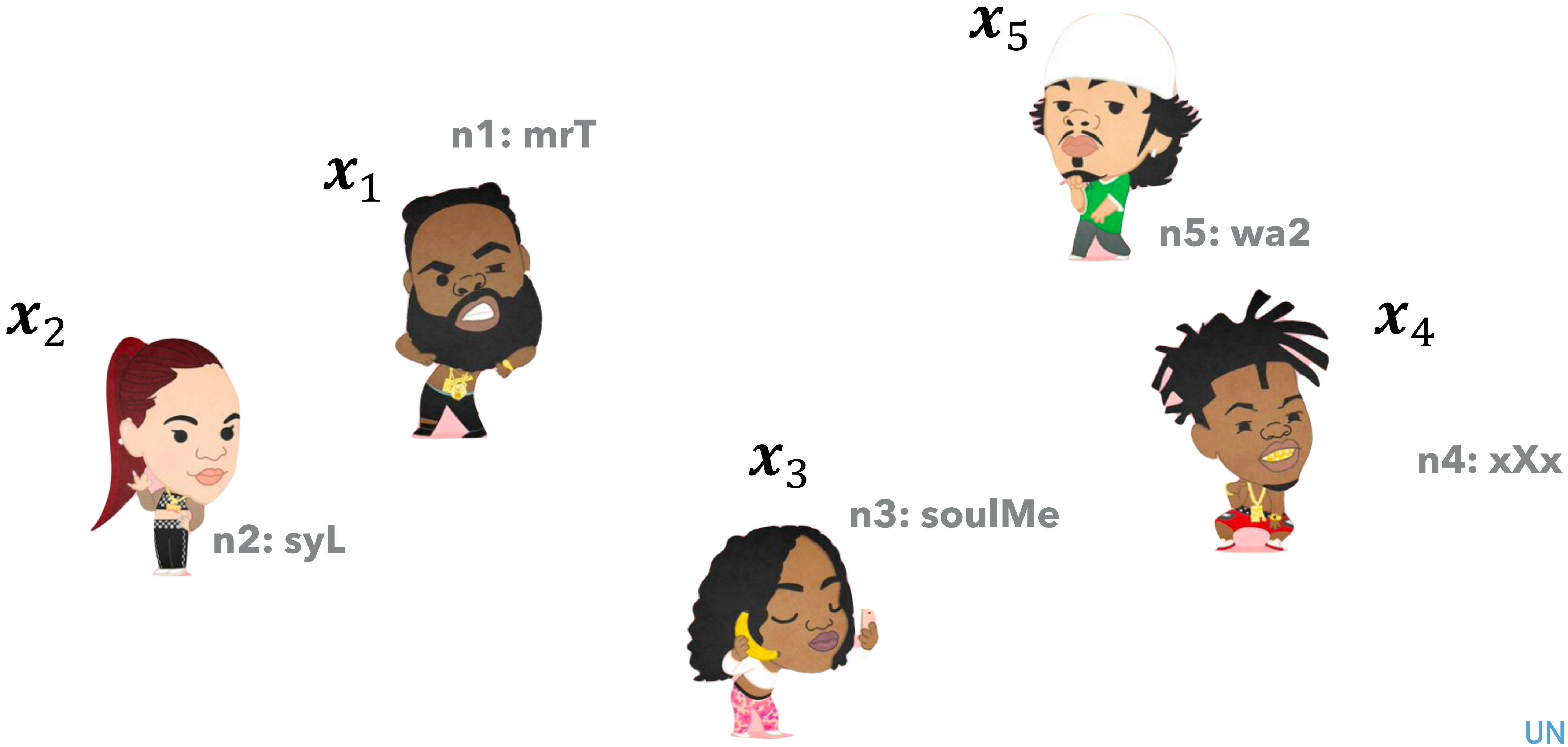
[in a principled, data-driven manner]

Example: Musical interaction

Wu-tang Clan ::Reborn::

$\mathcal{V}$ Nodes: Rappers

$\mathcal{Y}$ Measurements: [attributes] (features)

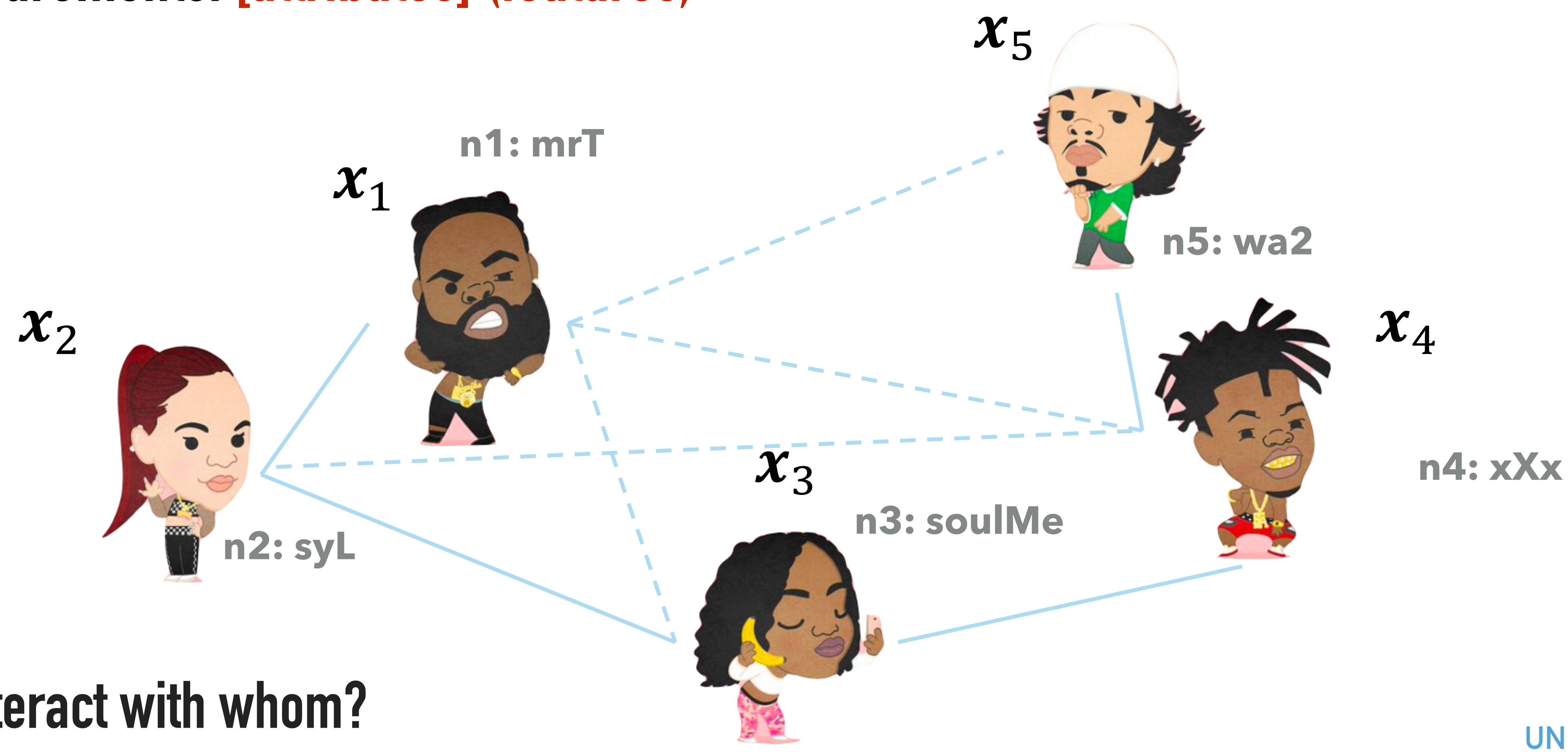


Example: Musical interaction

Wu-tang Clan ::Reborn::

$\mathcal{V}$ Nodes: Rappers

$\mathcal{Y}$ Measurements: [attributes] (features)



Who interact with whom?

Classical link learning

Task

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph. Given a set $\mathcal{E}_{\text{obs}} \subset \mathcal{E}$, and possibly various attributes $\{\boldsymbol{x}_i \in \mathbb{R}^d\}_{i=1}^{|\mathcal{V}|}$, learn $\mathcal{E}_{\text{miss}} := \mathcal{E} \setminus \mathcal{E}_{\text{obs}}$

Common Approaches

- Informal scoring methods [Liben-Nowell, et al., 03]
- (Partial) correlation networks [Efron, 07; Giannakis, 18]
- Regression-based methods [Hoff, 05]
- Graphical models [Dempster, 72; Meinshausen, 06; Kumar, 19]
- Hypothesis test methods [Drton, et al., 04]
- Graph signal processing methods [Kalofolias, 16; Dong, 16; Mateos 19]

How to extend this task to higher-order interactions?

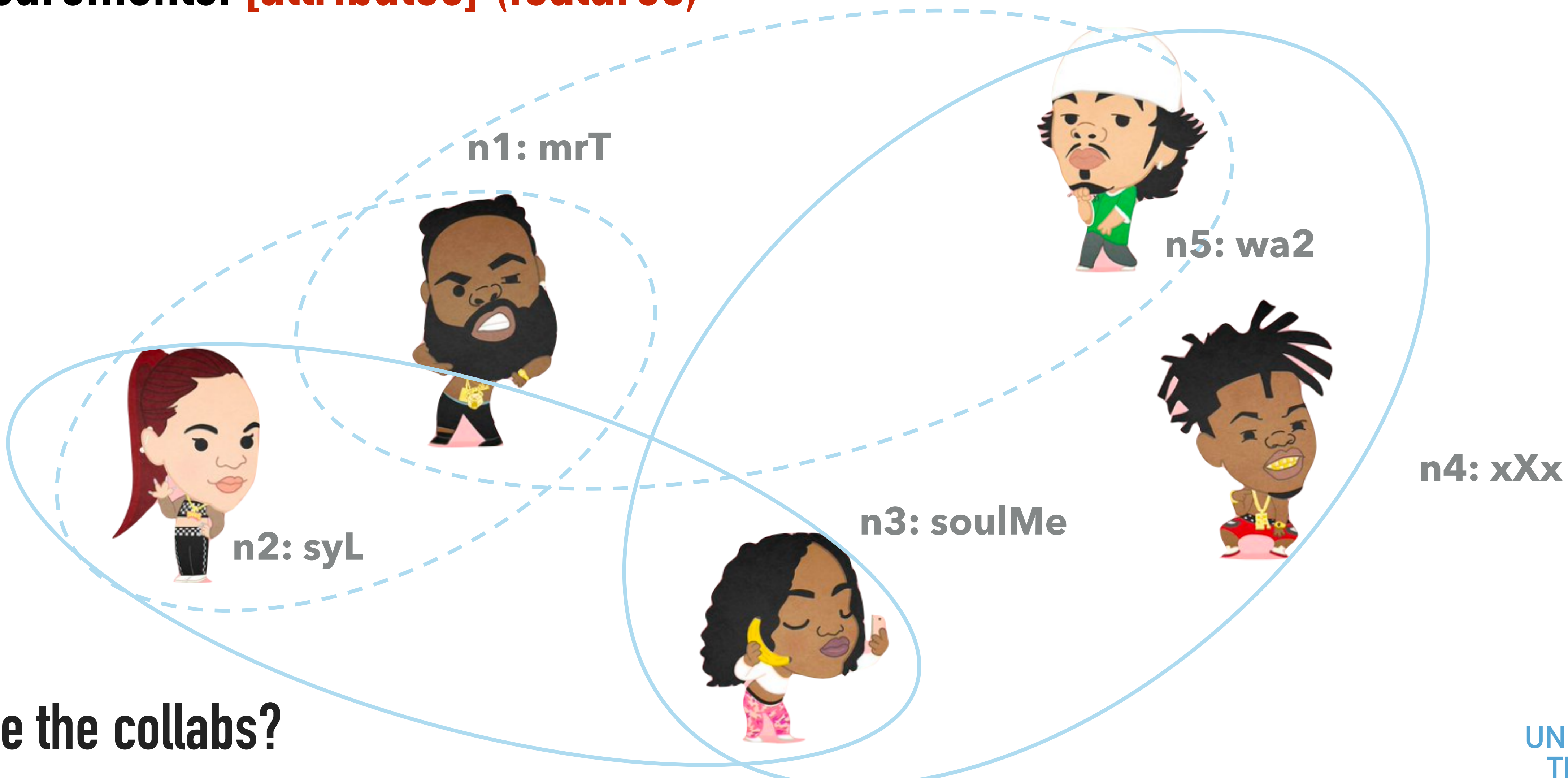
[in a principled, data-driven manner]

Example: Music collaboration

Wu-tang Clan ::Reborn::

$\mathcal{V}$ Nodes: Rappers

$\mathcal{Y}$ Measurements: [attributes] (features)



Which are the collabs?

Many ways to represent "higher-order" interactions

- **Hypergraphs** [Berge 89] :: edges join multiple nodes
- Set systems [Frank 95]
- **Tensors** [Kolda-Bader 09] :: tensor entries represent multinodal interactions
- Affiliation networks [Feld 81, Newman+ 02]
- Multipartite networks [Lind+ 07]
- **Abstract simplicial complexes** [Barbarossa 18] :: fully-connected subgraphs
- Multilayer networks [Kivela 89]
- Meta-paths [Sun-Han 12]
- **Projected representations** [Benson+ 15, 17] :: weighted graph representation

Representations of "higher-order" interactions

Common pitfalls of current models

- Make use of network structure directly, e.g., motif structures
- Physics-based assumptions that might not hold, e.g., flow assumptions

What is needed?

- Modeling tool based on the networked-data itself
- Expressibility to capture the role of the higher-order interactions
- Interpretability for predicting the appearance of higher-order relations

[Higher-order link learning in a principled, data-driven manner]

How to model then higher-order interactions?

Structural Equation Models

+ Successful accounting interactions
+ Model self-driven behaviors
+ Extended to capture nonlinearities

- Lack of higher-order link interpretability

$$x_i = \sum_{j \in \mathcal{V} \setminus i} a_{ij} x_j + \sum_{k \in \mathcal{V}} \gamma_{ik} \zeta_k$$

[J. Hox 98][X. Cai, 13][Giannakis 18]

Volterra Series

+ Widely-used for nonlinear dynamics
+ Captures complex dependencies
+ Theoretical guarantees

- Lack of self-driven relations

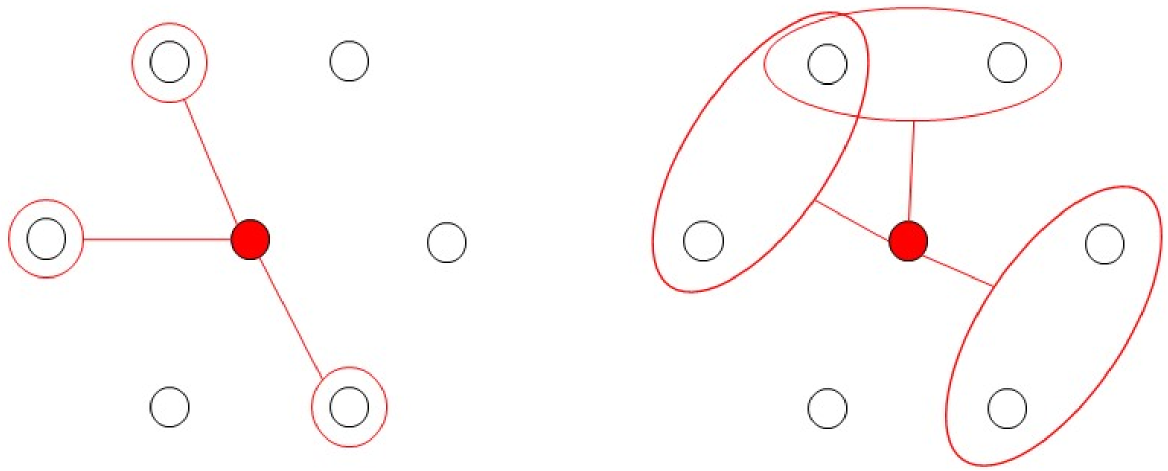
$$y(t) = h_0 + \sum_{p=1}^{P} \sum_{\tau_1 = a}^{b} \cdots \sum_{\tau_p = a}^{b} h_p(\tau_1, \ldots, \tau_p) \prod_{j=1}^{p} x(t - \tau_j)$$

[M. Schetzen, 80][V. Ketatos, 11]

Combining the best of both worlds…

Modeling higher-order interactions

Key idea:
Description of the ith nodal feature in terms of a set of subsets of nodes



Modeling higher-order interactions

Key idea:

Description of the ith nodal feature in terms of a set of subsets of nodes $\mathcal{S}_P^{(i)}$

$$\mathcal{S}_P^{(i)} := \bigcup_{p=1}^{P} \mathcal{S}_{*,p}^{(i)}, \; with \;\; \mathcal{S}_{*,p}^{(i)} := \bigcup_{l=1}^{L_p} \mathcal{S}_{l,p}^{(i)}$$

Set of subsets upto order P

denotes the lth set of p nodes related to the ith node in the graph

Nodal features $x_i = f\left(\mathbf{x}, \mathcal{S}_P^{(i)}\right), \forall\, i \in \{1, \ldots, N\}$ $\quad [\mathbf{x}]_i = x_i$

Nonlinear mapping

Self-driven graph Volterra models

Instantiation of model

Absorbs $\mathcal{S}_P^{(i)}$ set dependency

$$x_i = f_i(\mathbf{x})$$

Series Expansion

Order P expansion

$$f_i(\mathbf{x}) = h_o^{(i)} + \sum_{p=1}^{P} H_p^{(i)}[\mathbf{x}] + \epsilon_i$$

Expansion Module

Permutation invariant nonlinearity

$$H_p^{(i)}[\mathbf{x}] := \sum_{l=1}^{L_p} h_{l,p}^{(i)} \cdot g(\{x_q : q \in \mathcal{S}_{l,p}^{(i)}\})$$

[Coutino, et al, 20]

Self-driven graph Volterra models

Nodal measurement model

$$x_i = h_o^{(i)} + \sum_{p=1}^{P} H_p^{(i)}[\mathbf{x}] + \epsilon_i$$

As $\mathcal{S}_P^{(i)}$ it is usually unknown, we can expand the graph Volterra module

$$H_p^{(i)}[\mathbf{x}] = \sum_{k_1=1}^{N} \cdots \sum_{k_p=k_{p-1}}^{N} h_p^{(i)}(k_1, \dots, k_p) g(\{x_{k_q}\}_{q=1}^{p})$$

and associate the nonzero coefficients with the active sets…

[sparse coefficient expansion]
[for higher-order interaction discovery]

Now: [Identification]

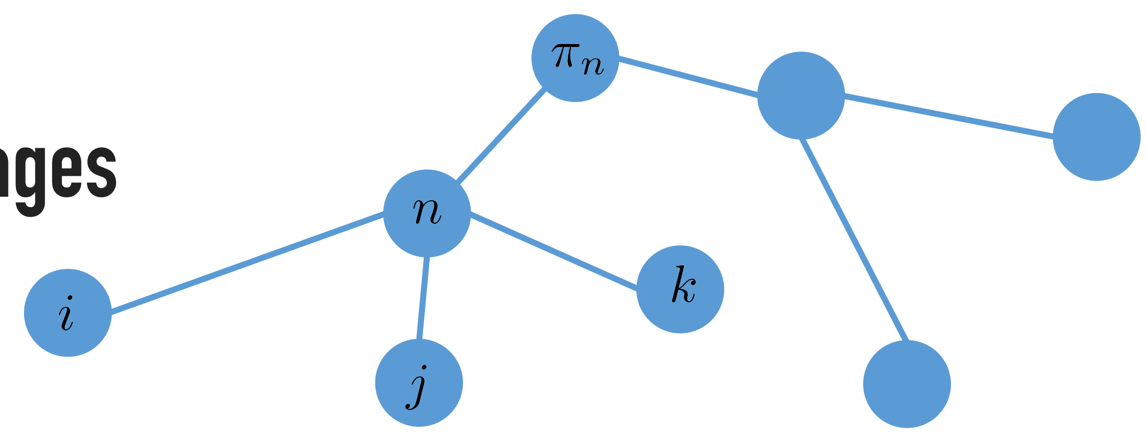
Topology identification from nodal attributes

[using self-driven graph Volterra models]

SD-GVM Application: Distribution Networks

Task:

Identify the network topology from nodal voltages



Key ideas:

Bus n voltage depends on its parent bus [Zhang, '13]

Bus n voltage also influences its children buses down the network [Yang, '20]

Model captures interactions between subset of buses

nth node parent

$$v_n = v_{\pi_n} + g_n(\{v_i\}_{i \in \mathcal{C}_n})$$

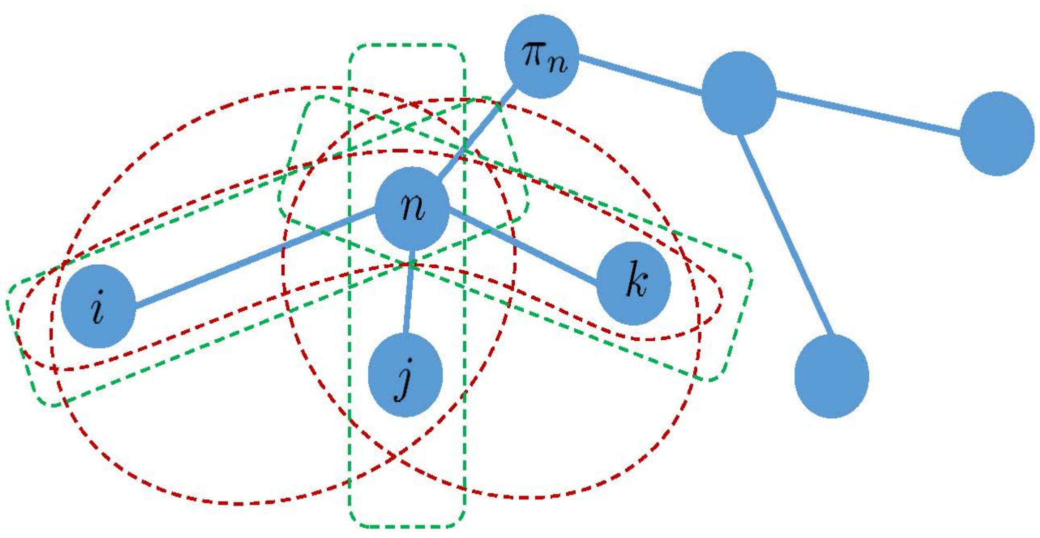
Nodal voltage nonlinear branch-flow model

nth node children

Linearization yields suboptimal performance: nonlinear models needed

SD-GVM Application: Distribution Networks

Interactions among nodal voltages



Nodal voltage nonlinear branch-flow model

$$v_n = v_{\pi_n} + g_n(\{v_i\}_{i\in\mathcal{C}_n})$$

Self-driven graph Volterra model

$$v_n = \sum_{i\in\mathcal{N}_0} \rho_i^{(n)} v_i + \sum_{i\in\mathcal{N}_0} \sum_{j\in\{k:k\in\mathcal{N}_0, k\geq i\}} \rho_{i,j}^{(n)} v_i v_j + \epsilon_n$$

First-order coeffs. ($\rho_i^{(n)}$) Second-order coeffs. ($\rho_{i,j}^{(n)}$)

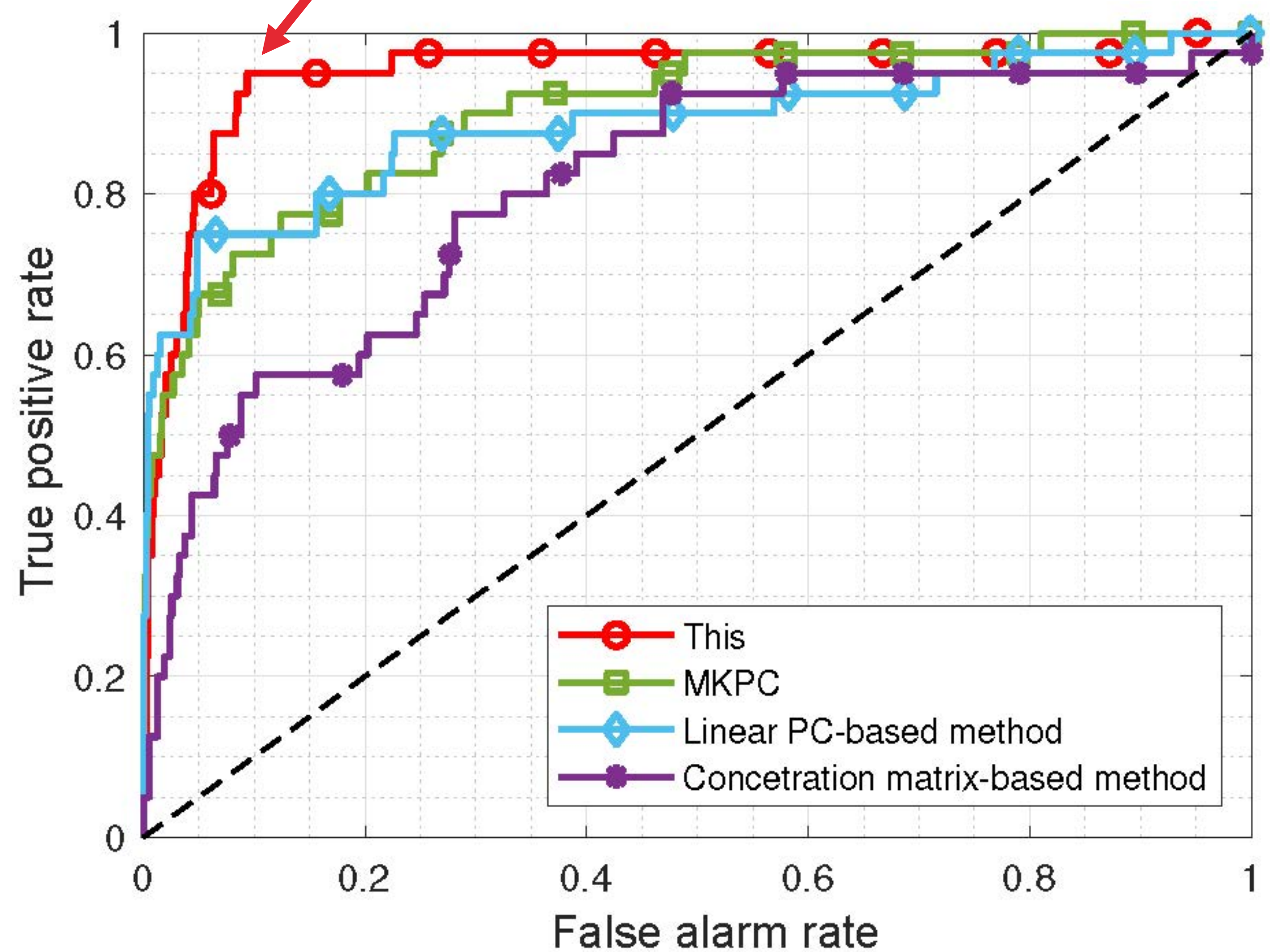
Nonzero coefficients capture interactions among pairs and triplets of nodal voltages enhancing the topology identification task [Yang, '20]

SD-GVM Application: Distribution Networks

Performance

- SCE 47-bus distribution grid
  - Real solar data from Smart* project
  - Voltage magnitudes from MATPOWER
- Baselines
  - Linear PC [Bolognani et al.'13]
  - Multi-kernel PC (MKPC) [Zhang et al.'17]
  - Concentration matrix [Deka et al.'17]

Higher-order interactions aware



[Yang et al.'20]

Now: [Prediction]

Predicting "group" interactions from previous interaction data

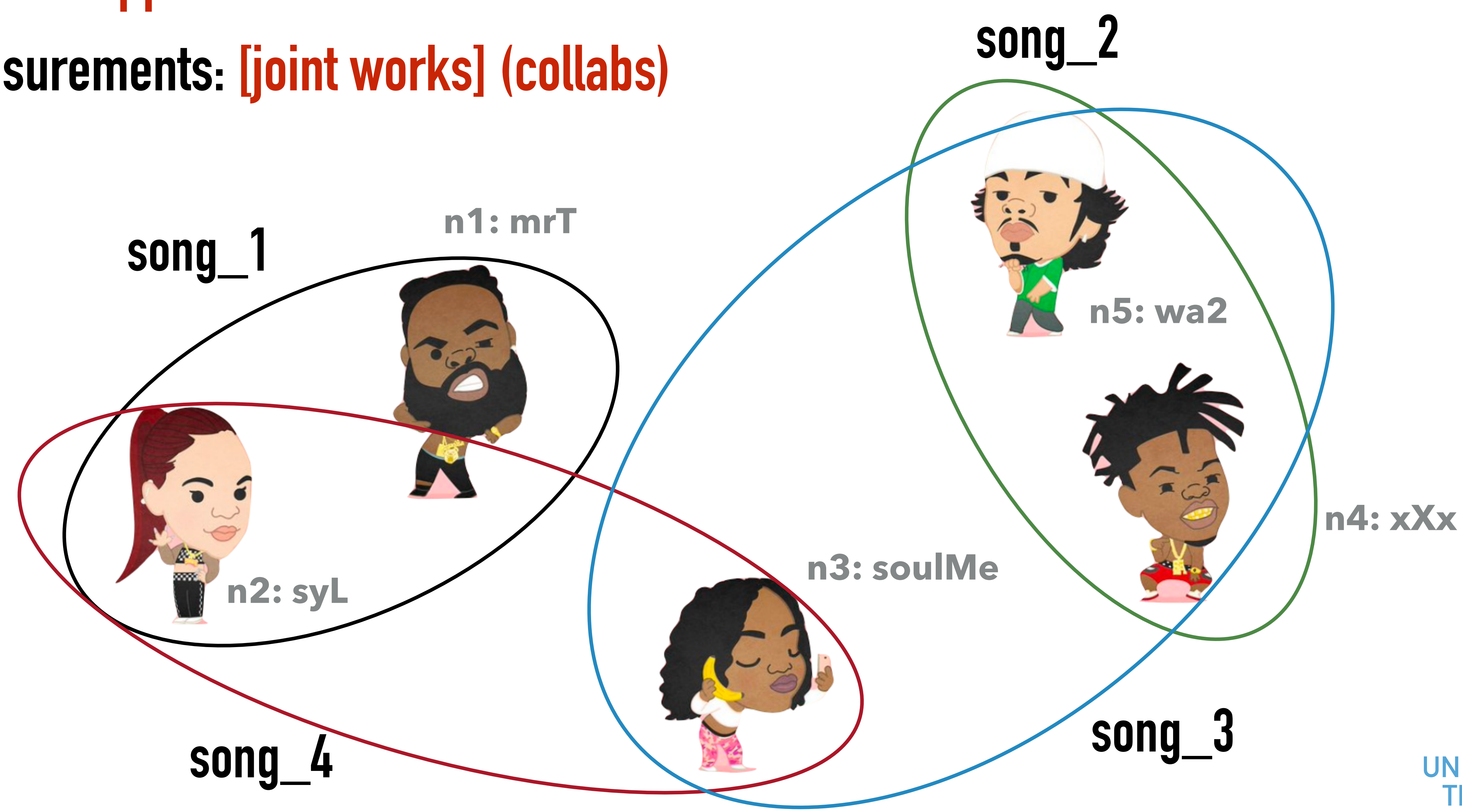
[using self-driven graph Volterra models]

Example: Music collaboration

Wu-tang Clan ::Reborn::

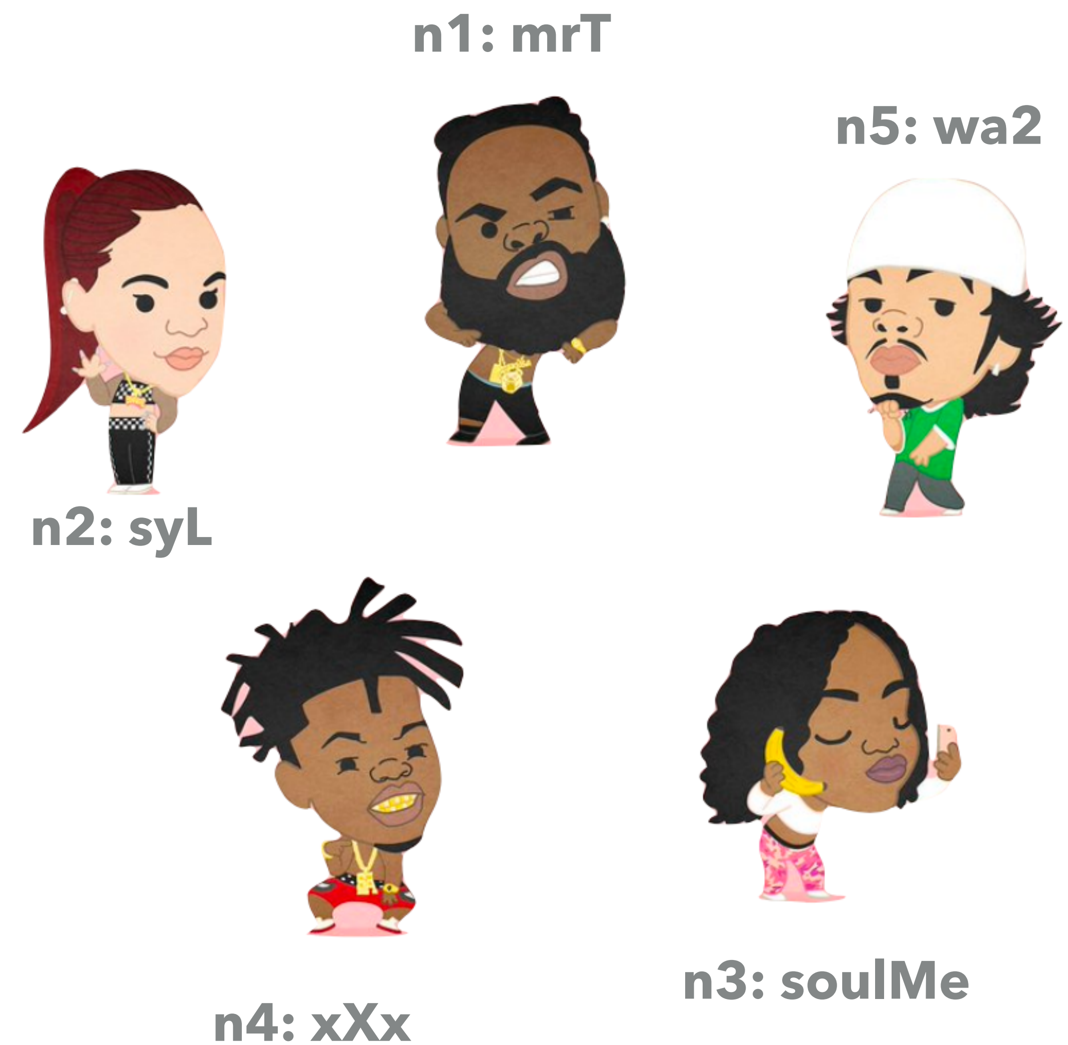
$\mathcal{V}$ Nodes: Rappers

$\mathcal{Y}$ Measurements: [joint works] (collabs)



Simplices as data

	Release	Authors	
t_1	song_1	{mrT, syL}	S_1
t_2	song_2	{wa2, xXx}	S_2
t_3	song_3	{wa2, xXx, soulMe}	S_3
t_4	song_4	{soulMe, syL}	S_4
.		. .. . . .	
t_5	song_5	???	



Simplex: set of observed nodes at a given time instant, e.g., S_t

Higher-order link (HO link) prediction: Who are writing the next songs?

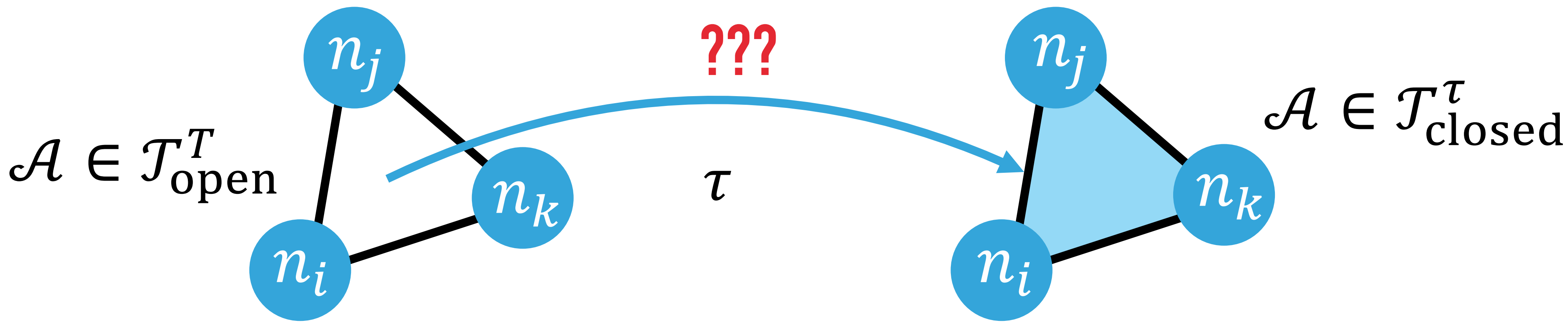
Simplified HO link prediction

Task:

Given $\{S_t\}_{t=1}^{T}$ **, and a set** $\mathcal{A} \subset \mathcal{T}_{\text{open}}^{T}$ **, predict if** $\exists \tau, \tau > T: \mathcal{A} \subset \mathcal{T}_{\text{closed}}^{\tau}$

$$\mathcal{T}_{\text{open}}^{T} = \{open\ triangles\ upto\ time\ T\}$$

$$\mathcal{T}_{\text{closed}}^{T} = \{closed\ triangles\ upto\ time\ T\}$$



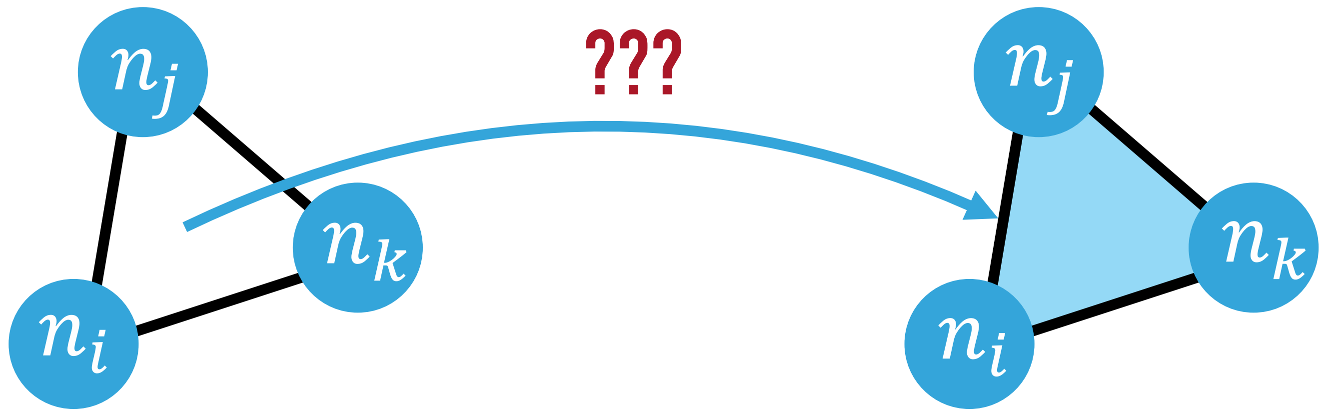
"From historical data, predict if an open triangle becomes closed."

still challenging, but more realistic

Triangle-link prediction: Scoring-based method

Task:

Given $\{S_t\}_{t=1}^{T}$, and a set $\mathcal{A} \subset \mathcal{T}_{\text{open}}^{T}$, predict if $\exists\tau, \tau > T: \mathcal{A} \subseteq \mathcal{T}_{\text{closed}}^{\tau}$



Approach:

1. compute scores for each open triangle
2. rank triangles by score
3. select highest scores as candidates.

many options for score function

$$s(i,j,k) = f(\{i,j,k\})$$

score

Scoring: Projected representation

Interactions graph

$\mathbf{A} \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|}$

Projected graph

$\mathbf{W} \in \mathbb{R}_+^{|\mathcal{V}|\times|\mathcal{V}|}$

Simplices

$t_1: \{1,2,3,4\}$

$t_2: \{1,3,5\}$

$t_3: \{1,6\}$

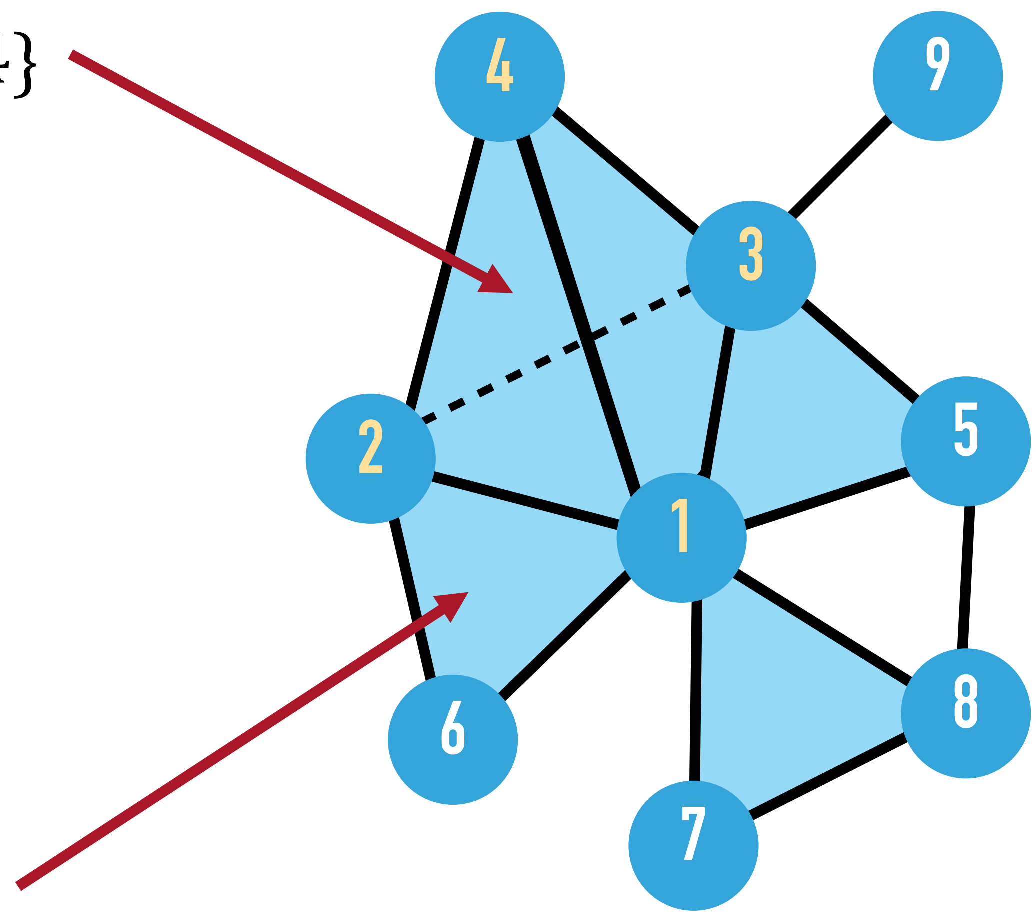
$t_4: \{2,6\}$

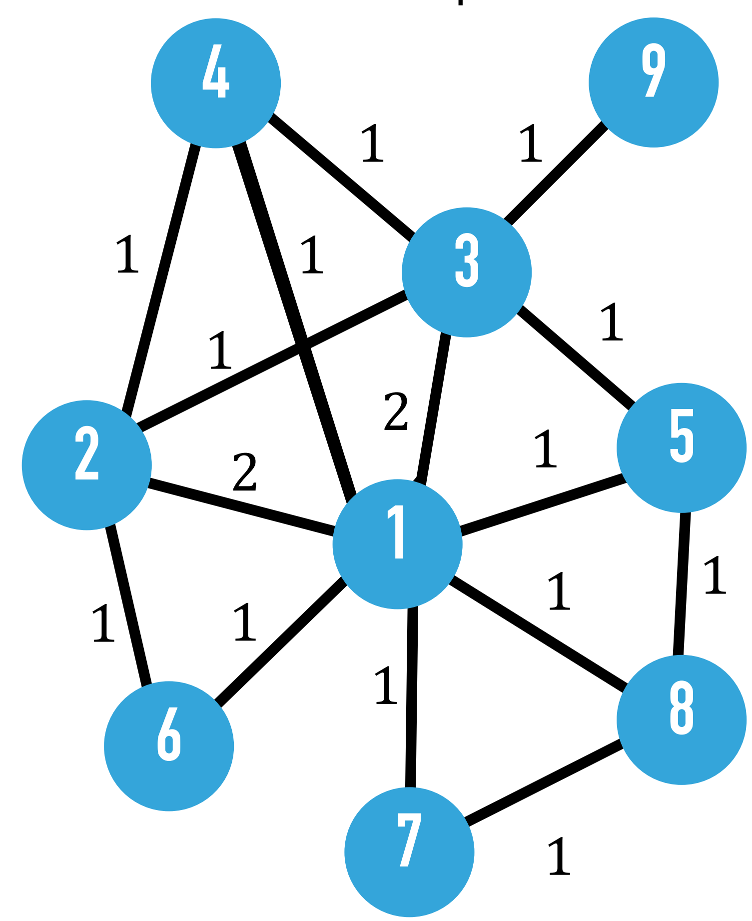
$t_5: \{1,7,8\}$

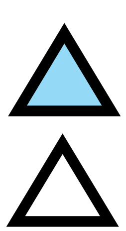
$t_6: \{3,9\}$

$t_7: \{5,8\}$

$t_8: \{1,2,6\}$







Closed **triangle**

Open **triangle**

Initial Connectivity

Weighted projected graph from 'filled structures'

$$[\mathbf{W}]_{ij} = |\{S_{t_k}: \{i,j\} \subset S_{t_k}\}|$$

[Benson , et al, 18]

Scoring: Common candidate functions

Possible score function candidates [Benson , et al, 18]

- Function of projected adjacency matrix, e.g.,

$$s(i,j,k) = ([\mathbf{W}]_{ij} * [\mathbf{W}]_{jk} * [\mathbf{W}]_{ik})^{1/3}$$

Geometric mean

- Function of one-hop neighbours, e.g.,

$$s(i,j,k) = \frac{|\mathcal{N}(i) \cap \mathcal{N}(j) \cap \mathcal{N}(k)|}{|\mathcal{N}(i) \cup \mathcal{N}(j) \cup \mathcal{N}(k)|}$$

Generalized Jaccard coefficient

- 'Global' similarity function, e.g.,

$$s(i,j,k) = \sum_{l,m \in \{i,j,k\}} [\mathbf{S}]_{lm}$$

PageRank

$$\mathbf{S} := (\mathbf{I} - \alpha \mathbf{W} \mathbf{D}_{\mathbf{W}}^{-1})^{-1}$$

- Learned function

$$s(i,j,k) = g_y(\mathbf{W}, \{i,j,k\})$$

ML Approach

Can we use self-driven Volterra models to define scoring functions?

[linking existence of link with graph Volterra kernels]

SD-GVM Application: Triangle-link prediction

graph Volterra coefficients
(constraints are included)

Let us consider the latent-variable model

$$z_i(t) = h_{o,i} + \mathbf{h}_{i,1}^T \mathbf{s}_t + \mathbf{h}_{i,2}^T (\mathbf{s}_t \boxtimes \mathbf{s}_t),$$

to model the probability

sigmoid function

ith node active

$$P([\mathbf{s}_t]_i = 1 | z_i(t)) = \sigma(z_i(t))$$

where $\mathbf{s}_t \equiv S_t$

[binary N-dimensional representation of a simplex]

model can be fitted using logistic regression techniques

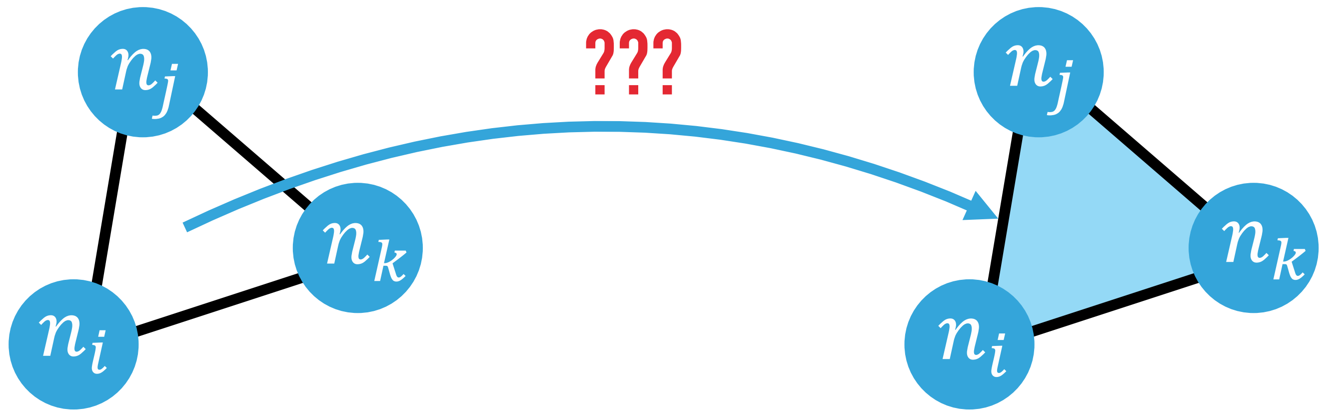
$$g(\mathcal{A}) = \prod_{a \in \mathcal{A}} a$$

$$\mathbf{a} \boxtimes \mathbf{a} := [a_1^2, a_1 a_2, \ldots, a_{N-1} a_N, a_N^2]^T$$

SD-GVM Application: Triangle-link prediction

Task:

Given $\{S_t\}_{t=1}^{T}$, and a set $\mathcal{A} \subset \mathcal{T}_{\text{open}}^{T}$, predict if $\exists \tau, \tau > T: \mathcal{A} \subseteq \mathcal{T}_{\text{closed}}^{\tau}$



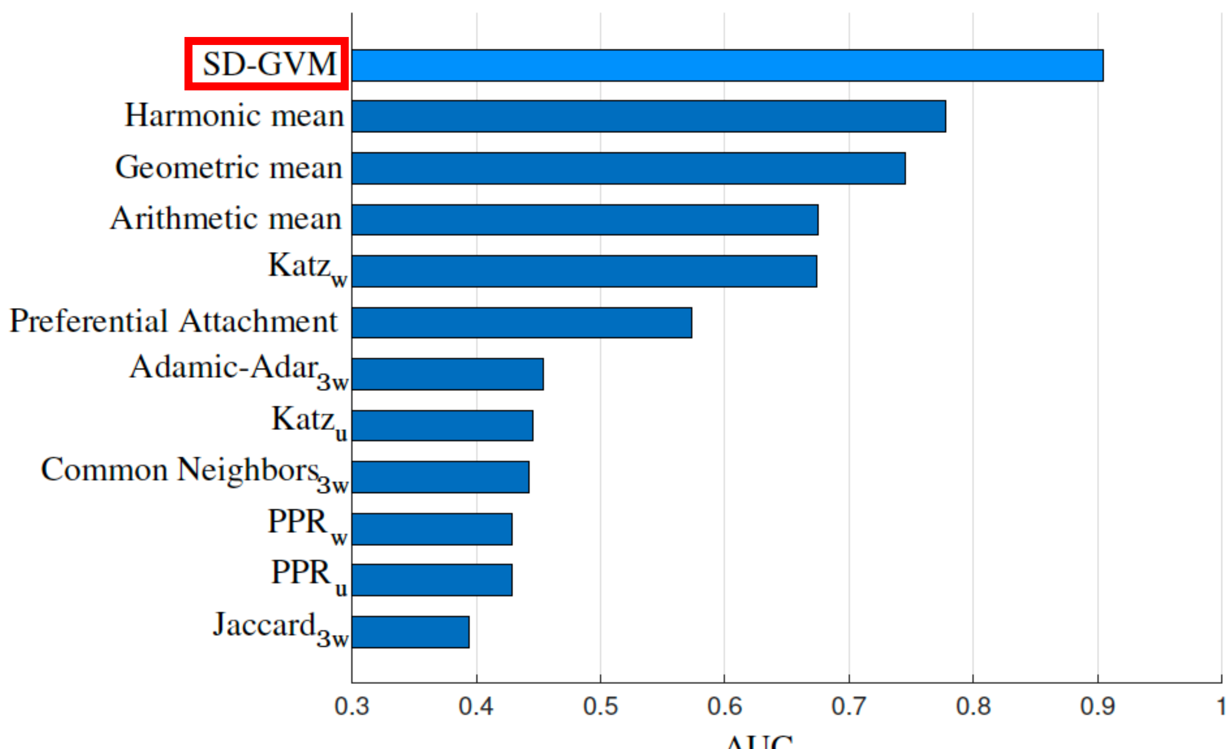
Approach:

1. From initial connectivity, find nonzero SD-GVM coefficients candidates(open triangles)
2. Fit the logistic regression SG-GVM to the simplex data
3. Select triads with highest absolute value coefficients as candidates

working assumption: $\left|h_2^i(j,k)\right| \propto p(\{i,j,k\} \subseteq \mathcal{T}_{\text{closed}}^{\tau} | \{S_t\}_{t=1}^{T})$

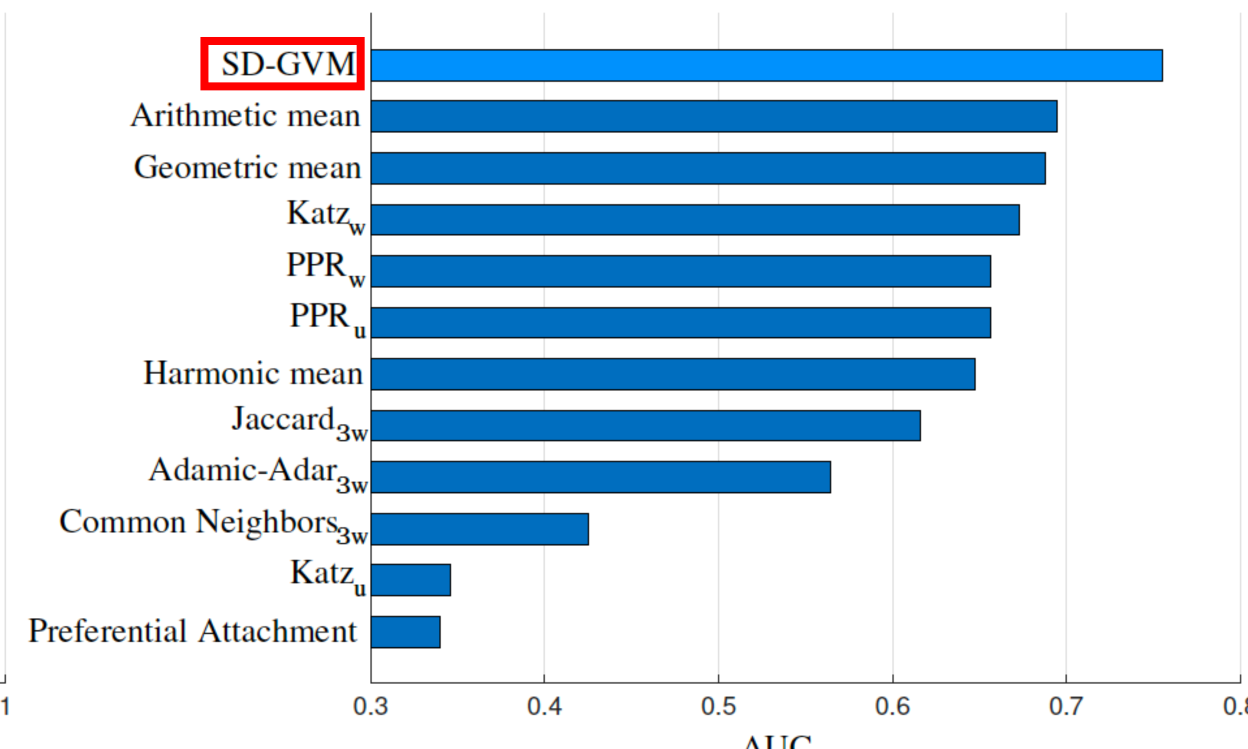
Application: Triangle-link prediction

Performance



(a)

Enron Emails dataset



(b)

Primary school dataset

[Coutino et al., 20]

Open Venues

[Montanari, et al. 16]

- Symmetric tensor completion for Triangle-link prediction

$$\mathcal{X}(i,j,k) = |\{S_{t_k}: \{i,j,k\} \subset S_{t_k}\}| \qquad \mathcal{X}(i,j,0) = |\{S_{t_k}: \{i,j\} \subset S_{t_k}\}|$$

$$\mathcal{X} \in \mathbb{R}_{+}^{(|\mathcal{V}|+1)\times|\mathcal{V}|\times|\mathcal{V}|}$$

- Adaptive diffusions for Triangle-link prediction [Berberidis, et al. 19]

$$s(i,j,k) = h(\mathbf{F}, \{i,j,k\}) \qquad \mathbf{F} := \sum_{l=0}^{L} \theta_l (\mathbf{W}\mathbf{D}_{\mathbf{W}}^{-1})^k$$

- Point process modelling…

Simplices as (correlated) point process

Simplices

t_1: {1,2,3,4}

t_2: {1,3,5}

t_3: {1,6}

t_4: {2,6}

t_5: {1,7,8}

t_6: {3,9}

t_7: {5,8}

t_8: {1,2,6}

Activation profile

$$\mathbf{S} = \begin{array}{c} n_1 \\ \\ \vdots \\ \\ \\ \\ \vdots \\ \\ n_9 \end{array} \overset{\begin{array}{cccccccc} S_{t_1} & & & \cdots & & & & S_{t_8} \end{array}}{\underset{\begin{array}{cccccccc} \mathbf{s}_{t_1} & \mathbf{s}_{t_2} & & \cdots & & & & \mathbf{s}_{t_8} \end{array}}{\begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}}}$$

$$\mathbf{W} = \mathrm{offdiag}(\mathbf{L}_S)$$

projected adj. matrix

$$\mathbf{L}_S := \mathbf{S}\mathbf{S}^T$$

Hypergraph incidence matrix

$$\mathbf{S} := [\mathbf{s}_{t_1}, \mathbf{s}_{t_1}, \dots, \mathbf{s}_{t_8}]$$

Wrapping up

- Pair-wise network modelling is not enough
  - higher-order interactions are missing
- Combination of SEMs and VSMs
  - expressibility and interpretability
- Including effects from HO interactions
  - benefits topology identification based on nodal attributes
- SD-GVMs are directly applicable to HO link prediction
  - however, there is a complexity challenge
- Incidence matrix representation of simplices
  - allows spatio-temporal point process-based analysis

Questions?

GEERT LEUS
DELFT UNIVERSITY OF TECHNOLOGY

LEARNING HIGHER-ORDER INTERACTIONS

WITH GRAPH VOLTERRA MODELS

Thanks!