UNIVERSITÀ
DEGLI STUDI
DI PADOVA

DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

Link Prediction on Real and Synthetic Complex Networks

Department of Information Engineering

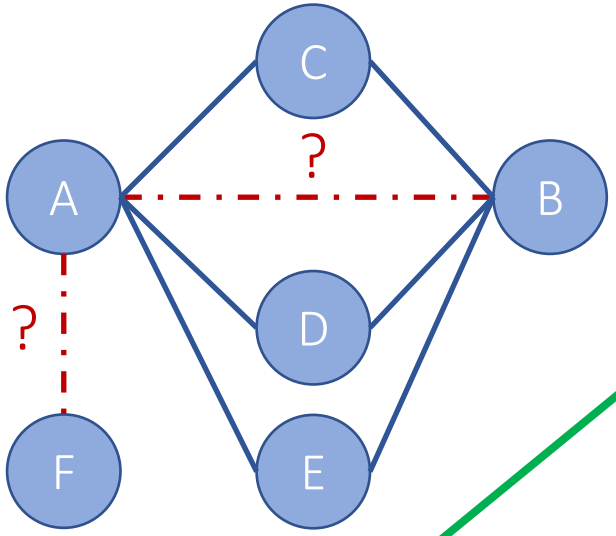
Master Candidate: *Umberto Michieli*

Supervisors: *Leonardo Badia (Università degli Studi di Padova)*

Carlo Cannistraci (Technische Universität Dresden)

10/09/2018

Topological Link Prediction (LP)



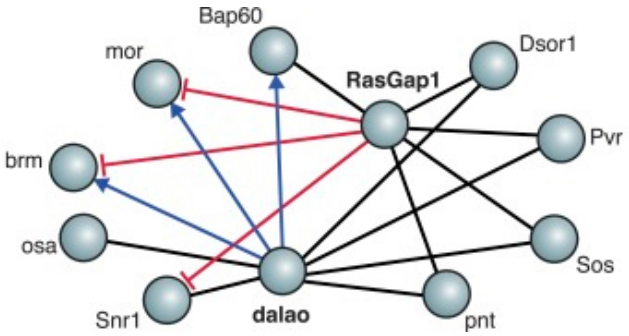
Problems:

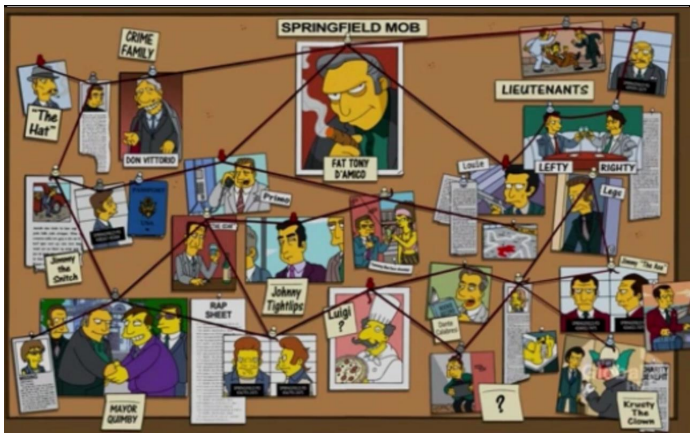
- Link forecast
- Partial information
- Reconstruction

Online social networks

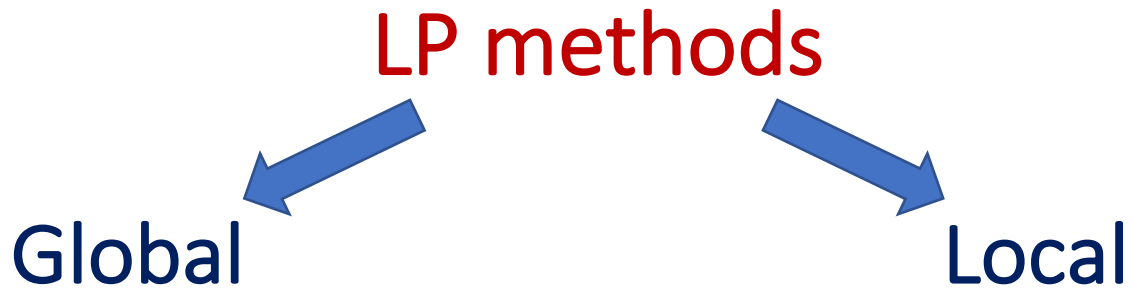
Biology

Covert networks





Motivation



- **Myth #1**: global methods are better
- **Myth #2**: SBM (global) should be the baseline
- No detailed LP test in the literature

→ Extensive LP evaluation

LP Methods

GLOBAL:

SPM Structural Perturbation Method [Lü et al. 2015]

SBM Stochastic Block Model (SBM) [Guimerà et al. 2009]

FBM Fast probability Block Model (FBM) [Liu et al. 2013]

DC SBM Degree Corrected SBM (DC SBM) [Karrer et al. 2011]

N SBM Nested SBM (N SBM) [Peixoto 2014]

DC N SBM DC and N SBM [Peixoto 2014]

LOCAL:

CH2-L2 Second variation of Cannistraci-Hebb on paths of length 2 (CH2-L2) [Muscoloni et al. 2018]

RA-L3 Resource Allocation on paths of length 3 (RA-L3) [Kovács et al. 2018]

Contribution

Standard procedure: remove 10% of links and compute likelihood scores

Mean precision, ranking and execution time

Real networks

- **Small**-size vs. **Large**-size

Synthetic networks

- **Hyperbolic geometry**: nonuniform Popularity-Similarity-Optimization (nPSO)
- **Euclidean geometry**: Watts-Strogatz (WS), Random Geometric Graph (RGG), Lancichinetti-Fortunato-Radicchi (LFR)

Contribution

Standard procedure: remove 10% of links and compute likelihood scores

Mean precision, ranking and execution time

Real networks

- **Small**-size vs. **Large**-size

Synthetic networks

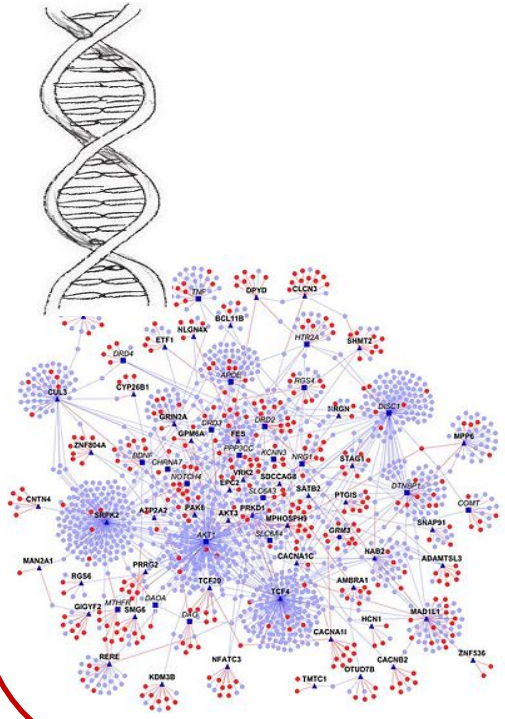
- **Hyperbolic geometry**: nonuniform Popularity-Similarity-Optimization (nPSO)
- **Euclidean geometry**: Watts-Strogatz (WS), Random Geometric Graph (RGG), Lancichinetti-Fortunato-Radicchi (LFR)

SBM-based!!

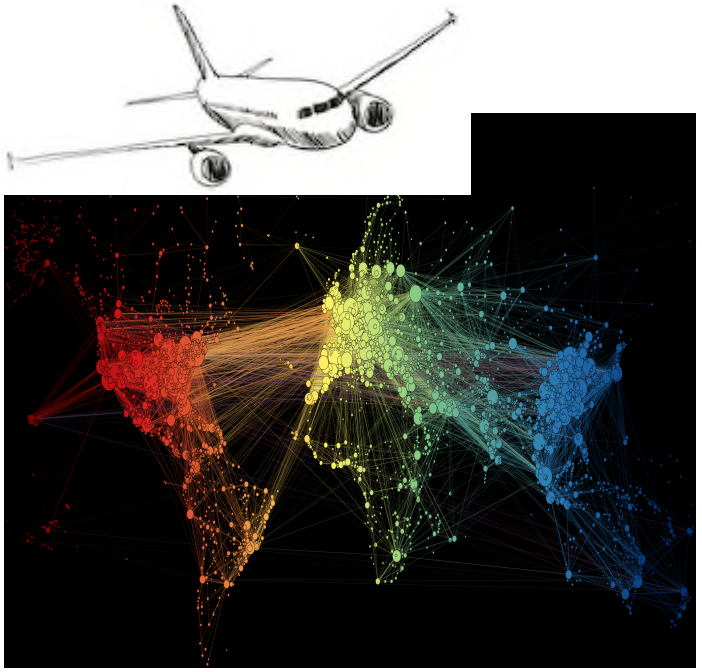
Small-size Real Networks

Networks of disparate fields of study

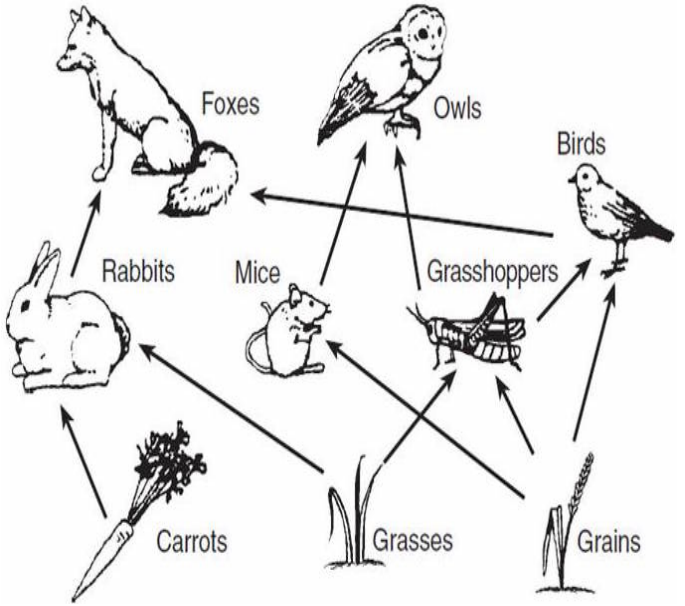
Biology



Transportation



Food-web



Small-size Real Networks

# networks	25
# nodes	10^1 - 10^3
avg. density	0.24
avg. power-law exponent (γ)	4.22

✗ non-hyperbolic

✗ non scale-free

	SPM	CH2-L2	SBM	FBM	RA-L3	SBM DC N	SBM DC	SBM N
Mean precision	**0.34**	0.30	0.28	0.27	0.26	0.22	0.21	0.06
Mean ranking	**2.1**	2.9	3.6	4.1	4.3	5.2	6.1	7.9
Mean time	sec	sec	hours	sec	sec	days	hours	days

Small-size Real Networks

# networks	25
# nodes	10^1 - 10^3
avg. density	0.24
avg. power-law exponent (γ)	4.22

✗ non-hyperbolic

✗ non scale-free

	SPM	CH2-L2	SBM	FBM	RA-L3	SBM DC N	SBM DC	SBM N
Mean precision	**0.34**	0.30	0.28	0.27	0.26	0.22	0.21	0.06
Mean ranking	**2.1**	2.9	3.6	4.1	4.3	5.2	6.1	7.9
Mean time	sec	sec	hours	sec	sec	days	hours	days

➢ Confirmed also on 486 structural connectomes (82 nodes)

Large-size Real Networks

Networks of disparate fields of study

Internet

Online Social Networks

Citation

Lexical

Large-size Real Networks

# networks	12
# nodes	10^3 to 10^4
avg. density	0.01
avg. power-law exponent (γ)	2.54

✓ hyperbolic

✓ scale-free

	CH2-L2	SPM
Mean precision	**0.19**	0.16
Mean ranking	**1.29**	1.71
Mean time	**0.9 h**	4.2 h

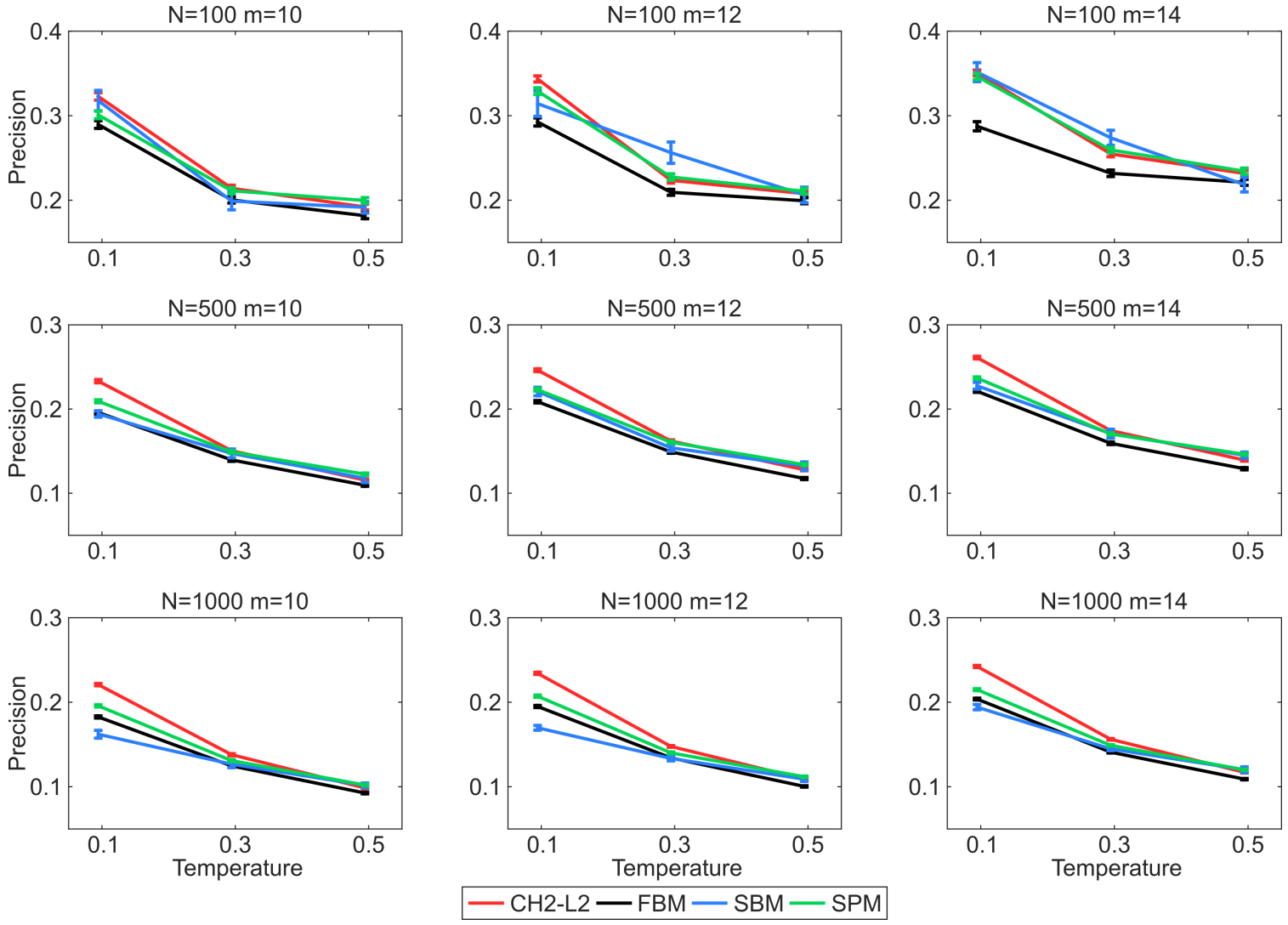
Hyperbolic Networks

- nPSO
  N, m, T,
  γ=3

scale-free

100 iterations

1) CH2-L2
2) SPM
3) SBM
4) FBM



Euclidean Networks (1/2)

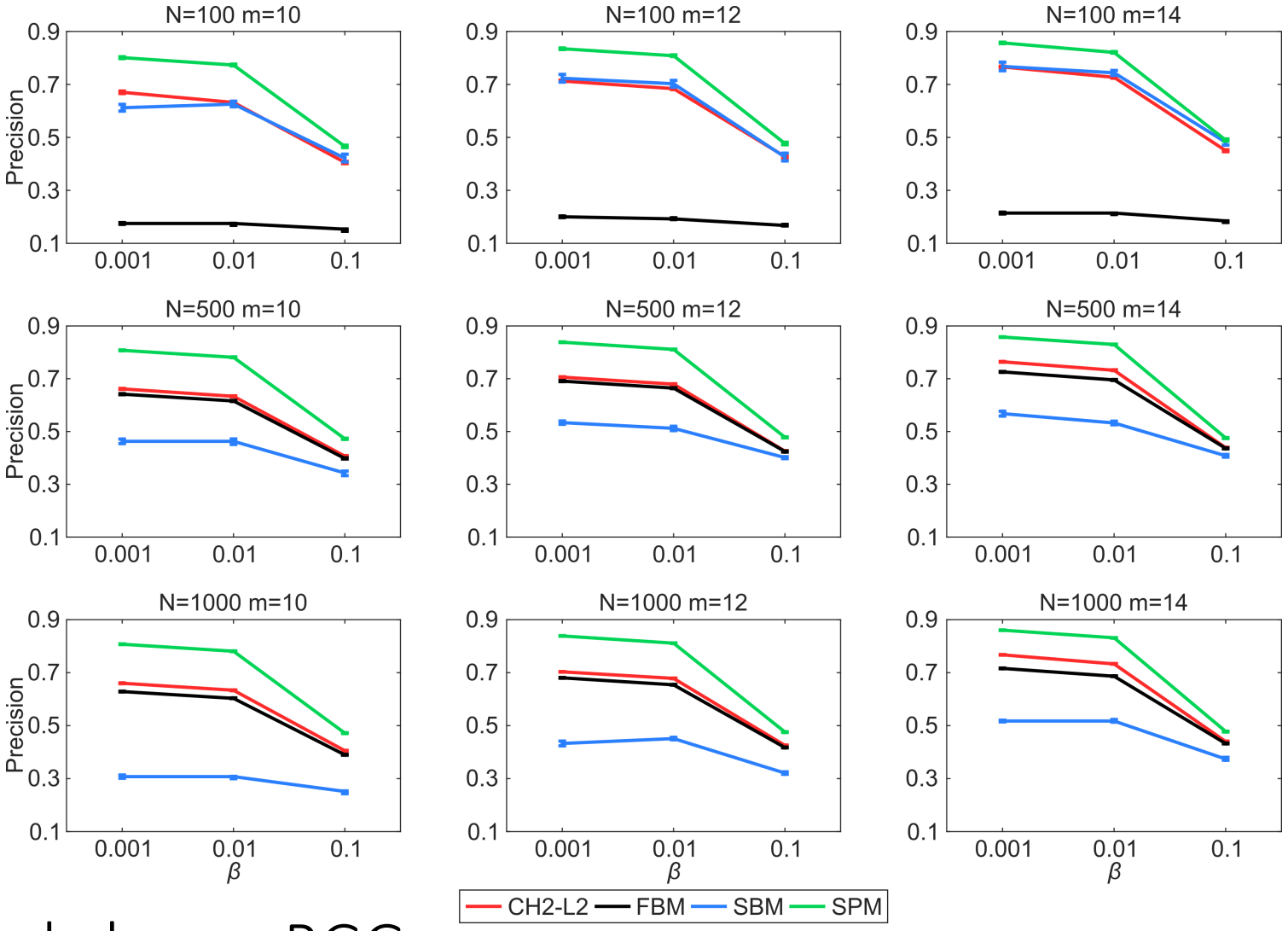
- WS

  N, m, β

non scale-free

100 iterations

1) SPM
2) CH2-L2
3) FBM
4) SBM



- Confirmed also on RGG

Euclidean Networks (2/2)

- **LFR**

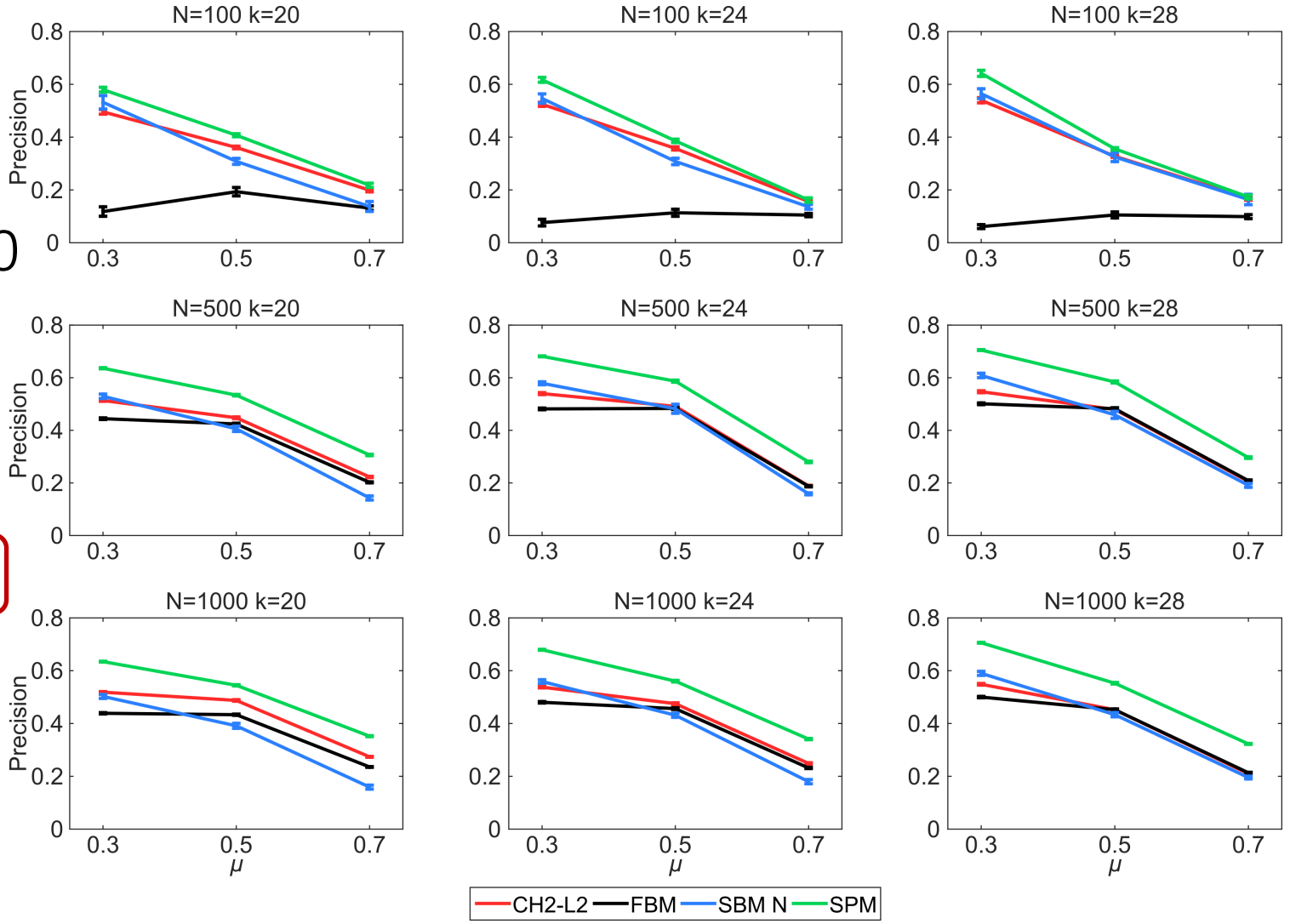
  N, m, μ,

  minc=N/10

scale-free

100 iterations

SBM-based



Euclidean Networks (2/2)

- **LFR**

  N, m, μ,

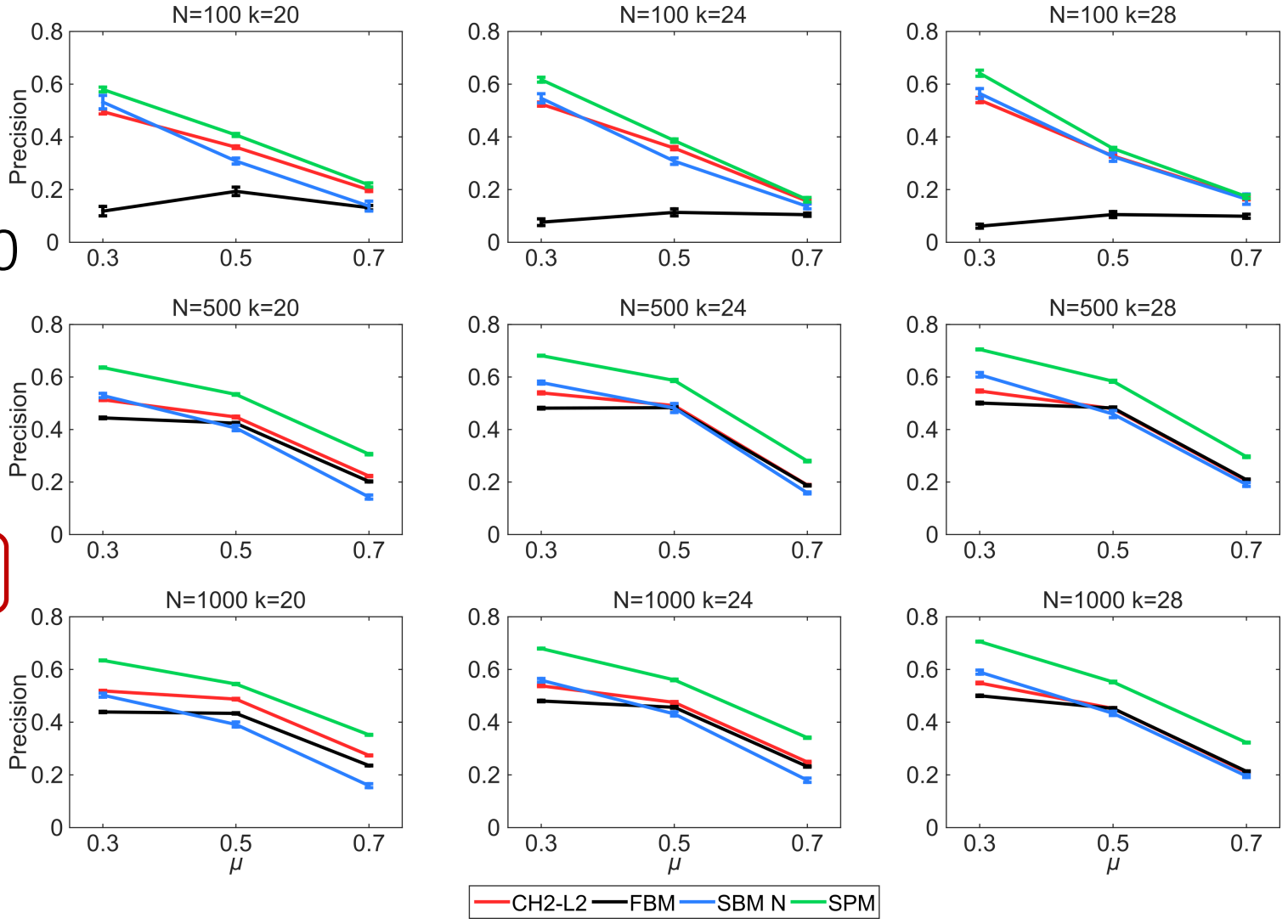
  minc=N/10

scale-free

100 iterations

SBM-based

1) SPM
2) CH2-L2
3) SBM N
4) FBM



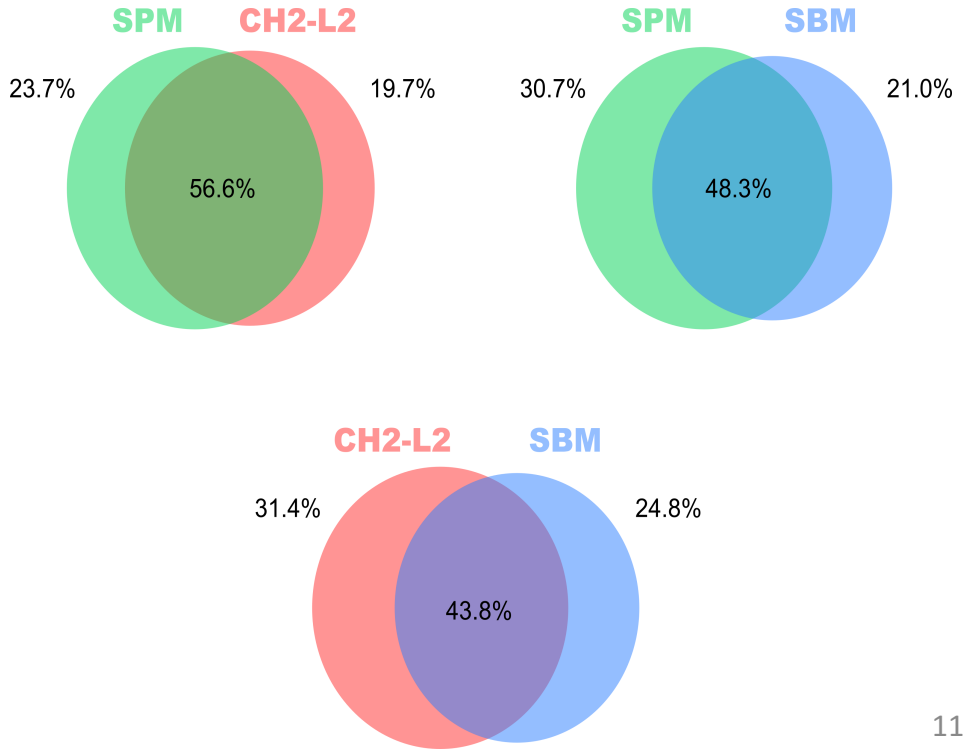
Conclusions

SPM & CH2-L2 are better baseline than SBM

Extensive LP test

Future Work

- Enlarge dataset
- CH2-L2, SPM and FBM on large-size networks
- Model with adjustable hyperbolicity
- Hybrid approach



Thank you for the attention!