Diversity vs Uniformity: Understanding the Evolution of Large Networks

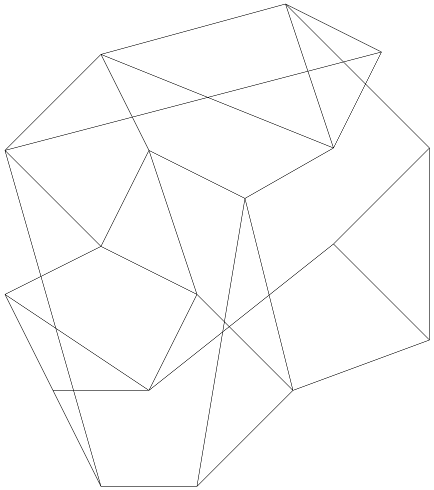
Jérôme Kunegis, UKOB
ROBUST, WP5

EC Project 257859

Diversity vs Regularity

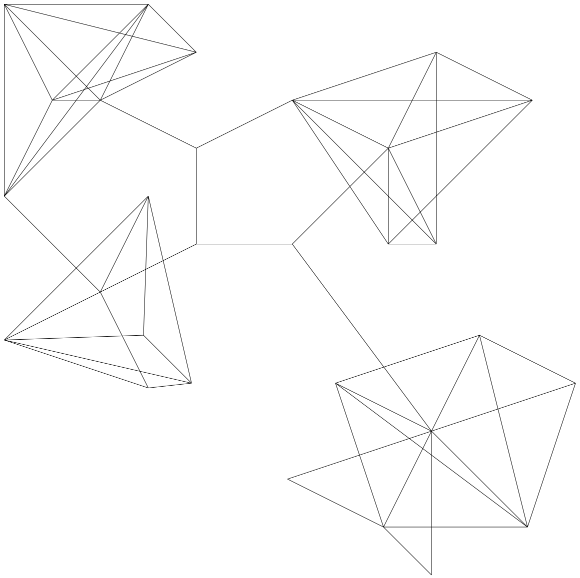
Diversity

- Many, equally-sized subcommunities
- High entropy
- ‘Flat’ structure

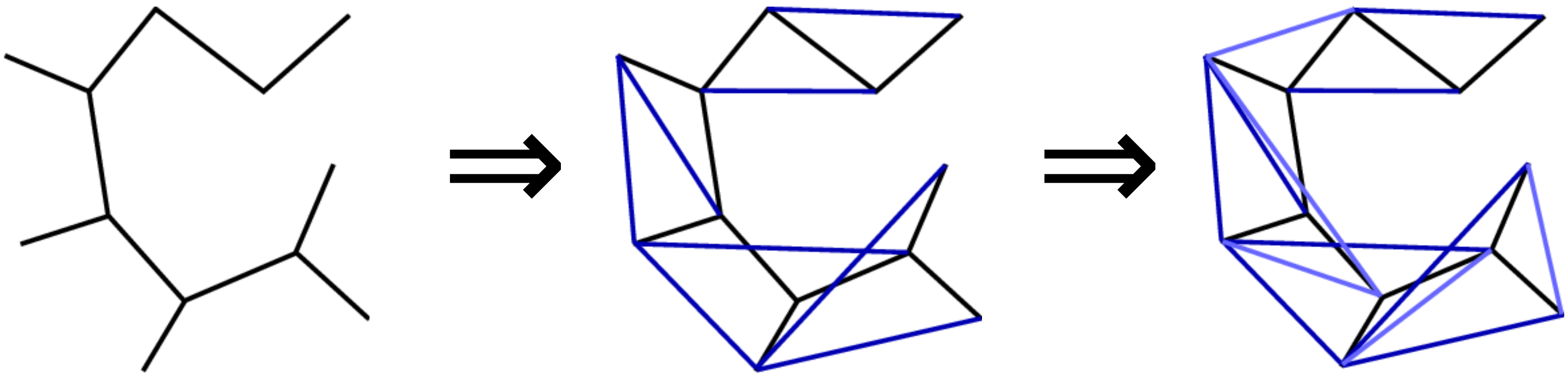


Regularity

- Few large subcommunities
- Low entropy
- Many ‘hubs’



Network Evolution



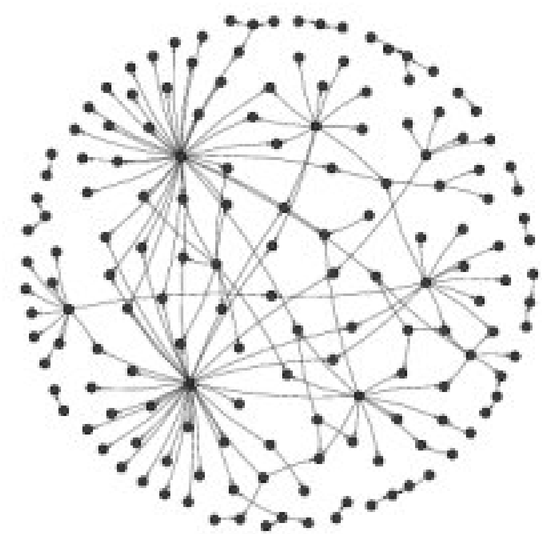
- How did a network look at time *t*?
- Idea: Observe the change of diversity/regularity over time

Outline

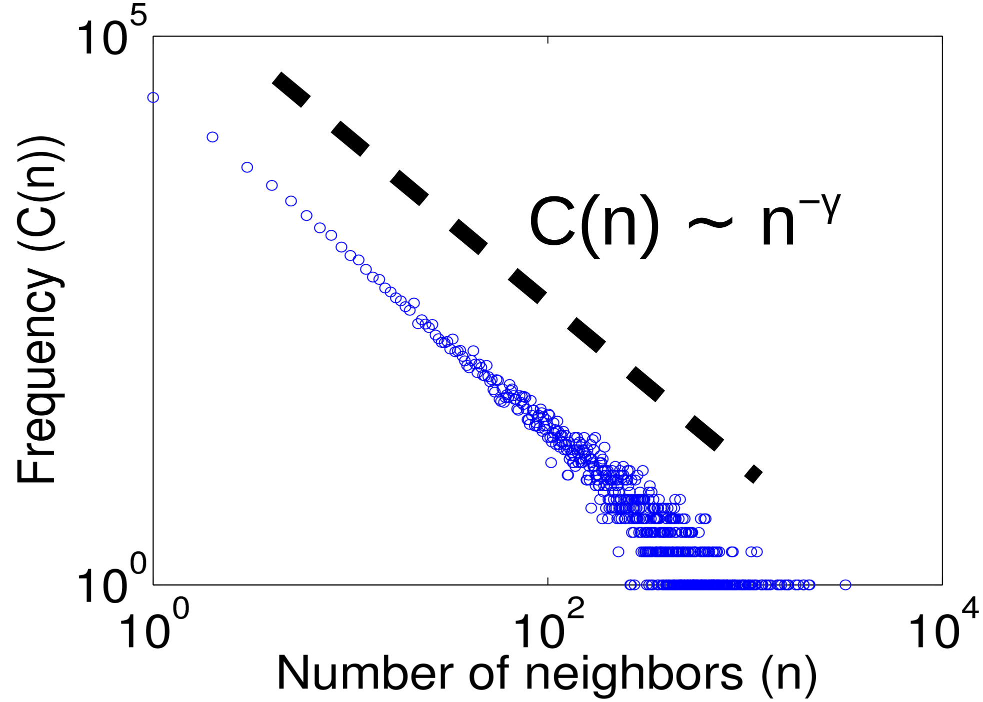
1. Power-law exponent
2. Weighted spectral distribution
3. Network entropy
4. Network rank

1. Power-law Exponent

Number of neighbors is unevenly distributed:



Epinions trust network (Massa et al. 2005)

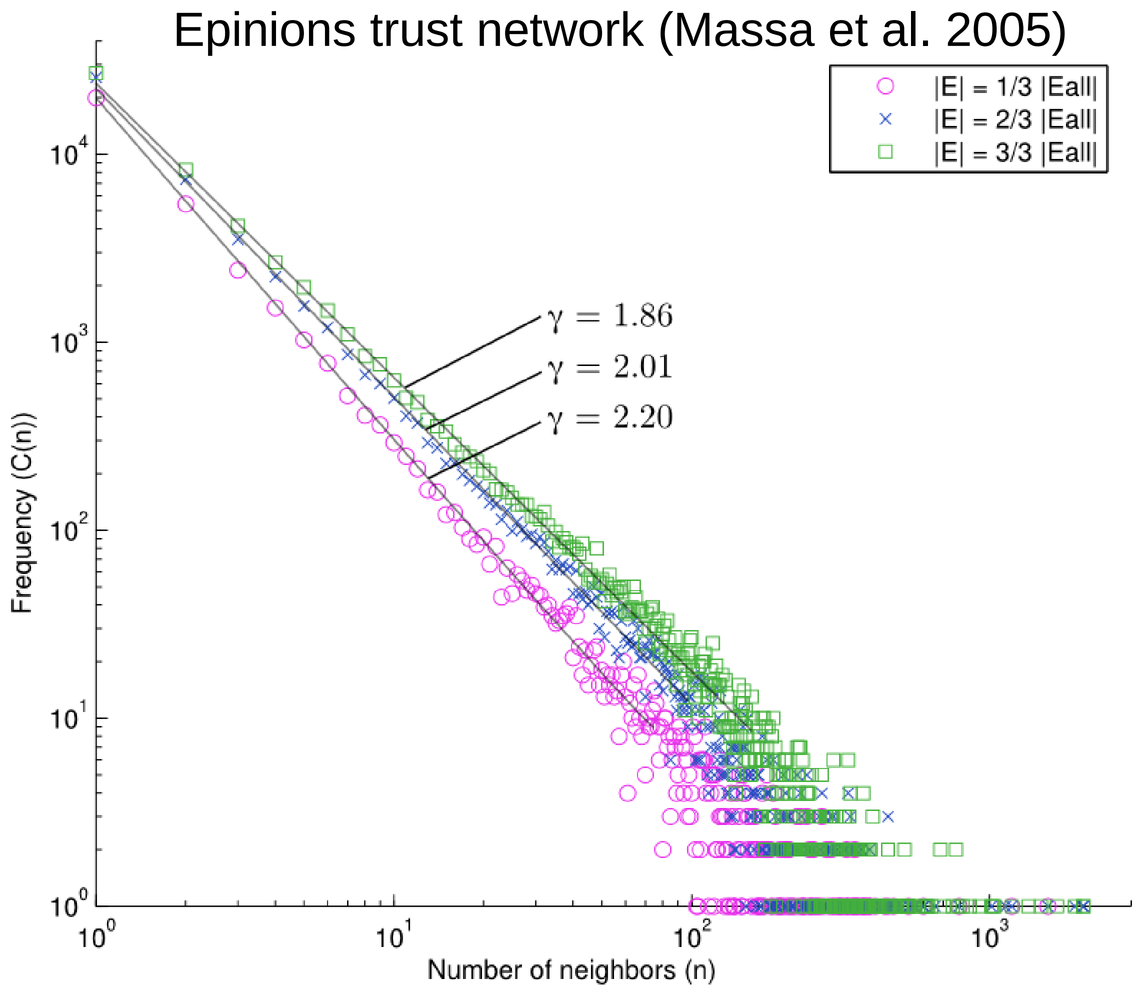


Results in a power-law (Newman 2006)

Exponent γ denotes regularity

1. Power-law Exponent over Time

ROBUST



γ shrinks ⇒ Network becomes more regular

WeST People and Knowledge Networks

2. Weighted Spectral Distribution

ROBUST

- Consider the $n \times n$ matrix $\mathbf{N}$ defined by

$\mathbf{N}_{ij} = 1 \ / \ \mathrm{sqrt}(d(i)d(j))$ when (i,j) is an edge

$\mathbf{N}_{ij} = 0$ otherwise

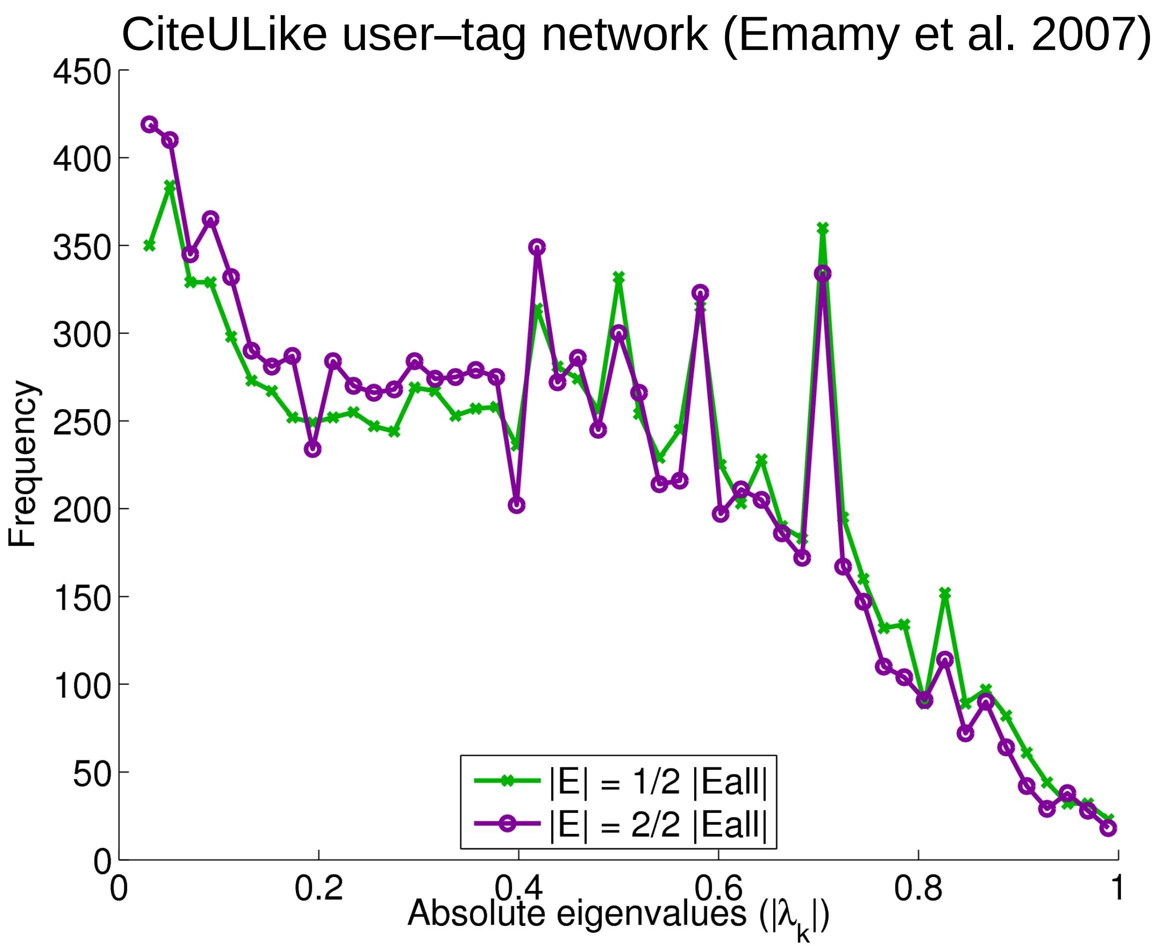
Then the distribution of the eigenvalues of $\mathbf{N}$ is called the weighted spectral distribution (WSD) (Fay et al. 2010)

Eigenvalues nearer to ± 1: diversity

Eigenvalues nearer to 0: regularity

2. Weighted Spectral Distribution over Time

ROBUST



The WSD shifts towards zero ⇒ The network becomes regular

WeST People and Knowledge Networks

3. Network Entropy

- Write the graph G as a sum of subgraphs G_k

$$G = G_1 \cup G_2 \cup \ldots \cup G_r$$

Each G_k has weighted edges, with total weight λ_k

- When picking an edge from G at random, the probability of it being in community G_k is

$$\lambda_k / (\lambda_1 + \lambda_2 + \ldots + \lambda_r) = \lambda_k / L$$

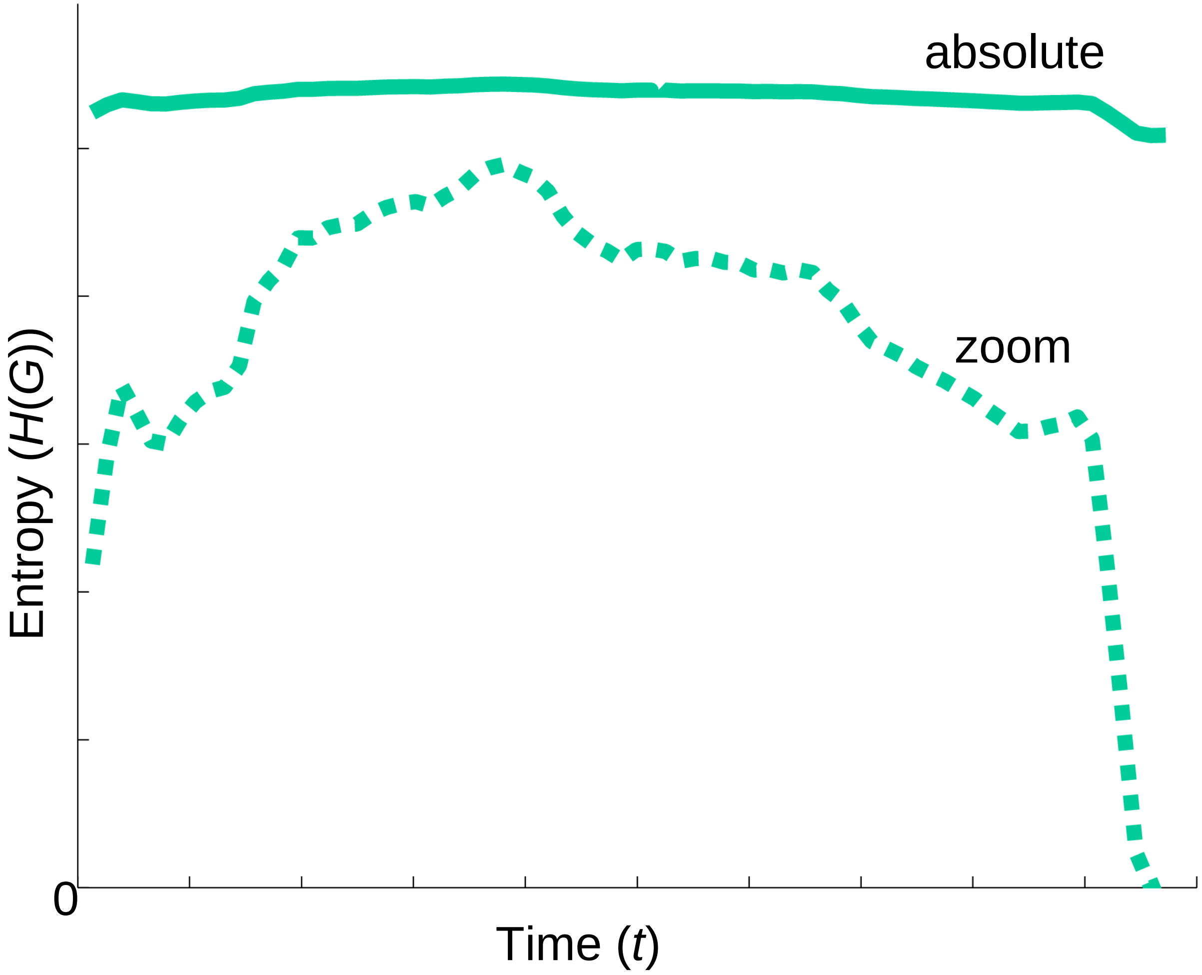
- The entropy of this distribution is (Kunegis et al. 2011)

$$H(G) = -\sum_k (\lambda_k / L) \log (\lambda_k / L)$$

- Entropy: Effective number of subcommunities

3. Network Entropy over Time

Enron email network (Klimt et al. 2004)



4. Network Rank

Decompose network into subcommunities:

$$G = G_1 \cup G_2 \cup \ldots \cup G_r$$

The rank r is a measure of diversity:

$$\operatorname{rank}(G) = r$$

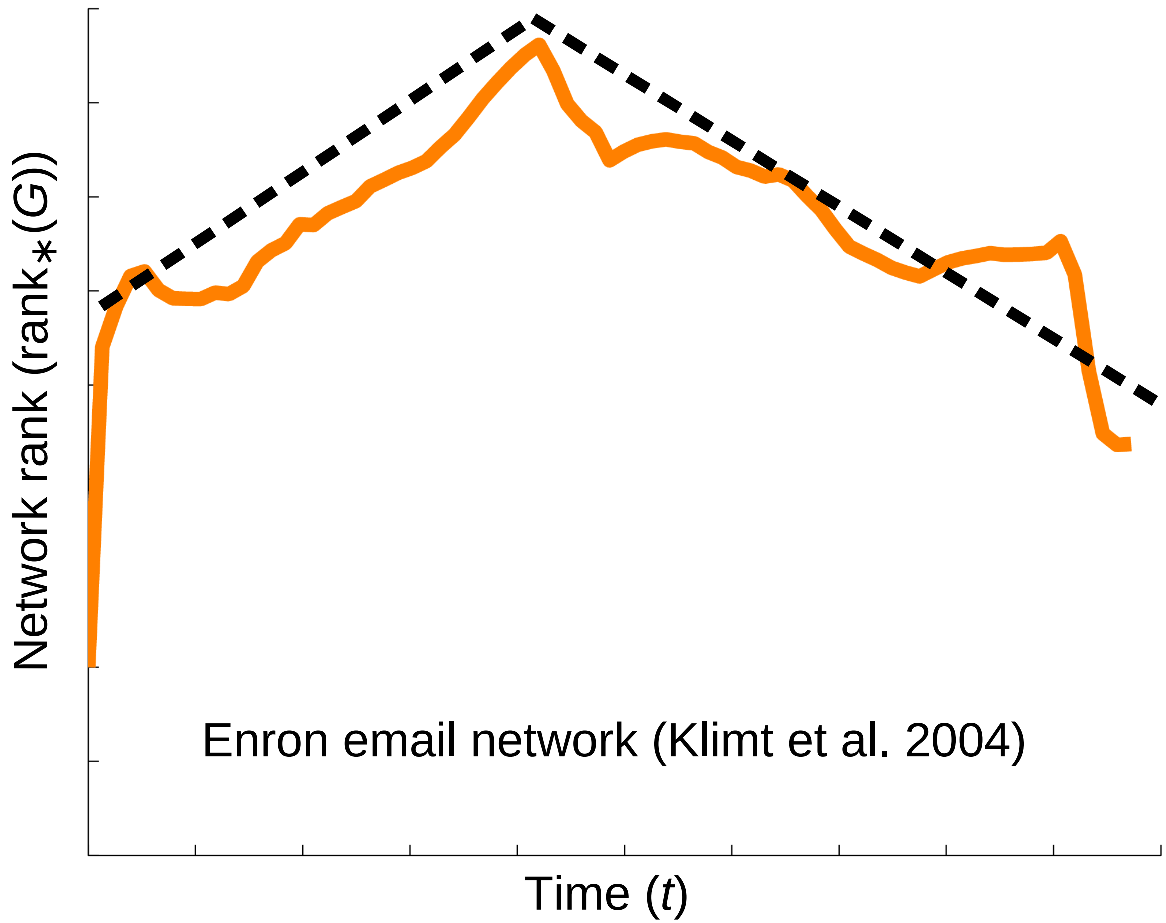
Weighted rank:

$$\operatorname{rank}_*(G) = \sum_k |G_k| \ / \ |G_1|$$

Robust measure of diversity (Kunegis et al. 2011)

4. Network Rank over Time

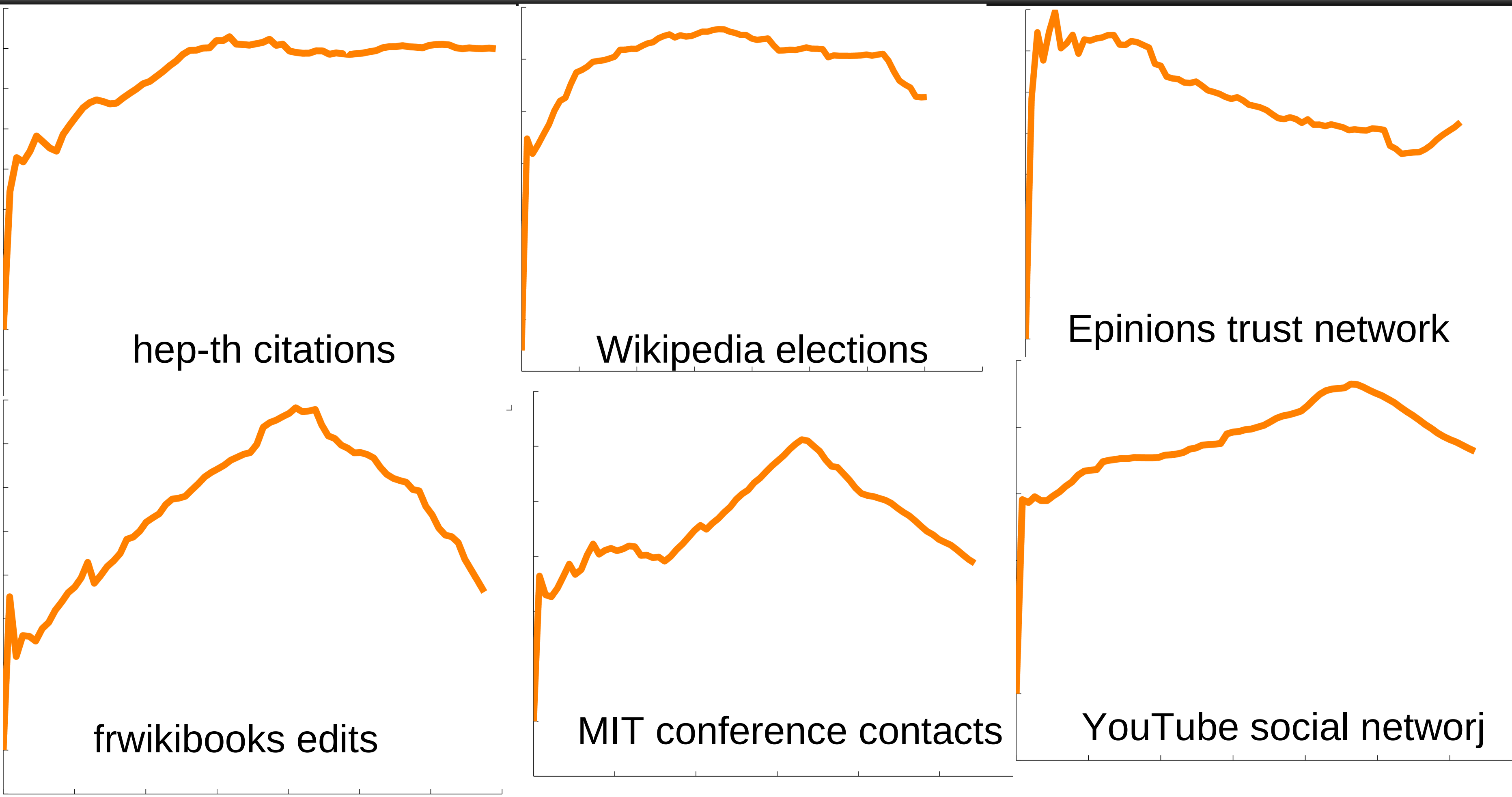
ROBUST



- Increasing network rank: increasing diversity
- Shrinking network rank: shrinking diversity

WeST People and Knowledge Networks

More Network Rank Plots



(biased towards good examples of convex evolution)

WeST People and Knowledge Networks

Conclusion

ROBUST

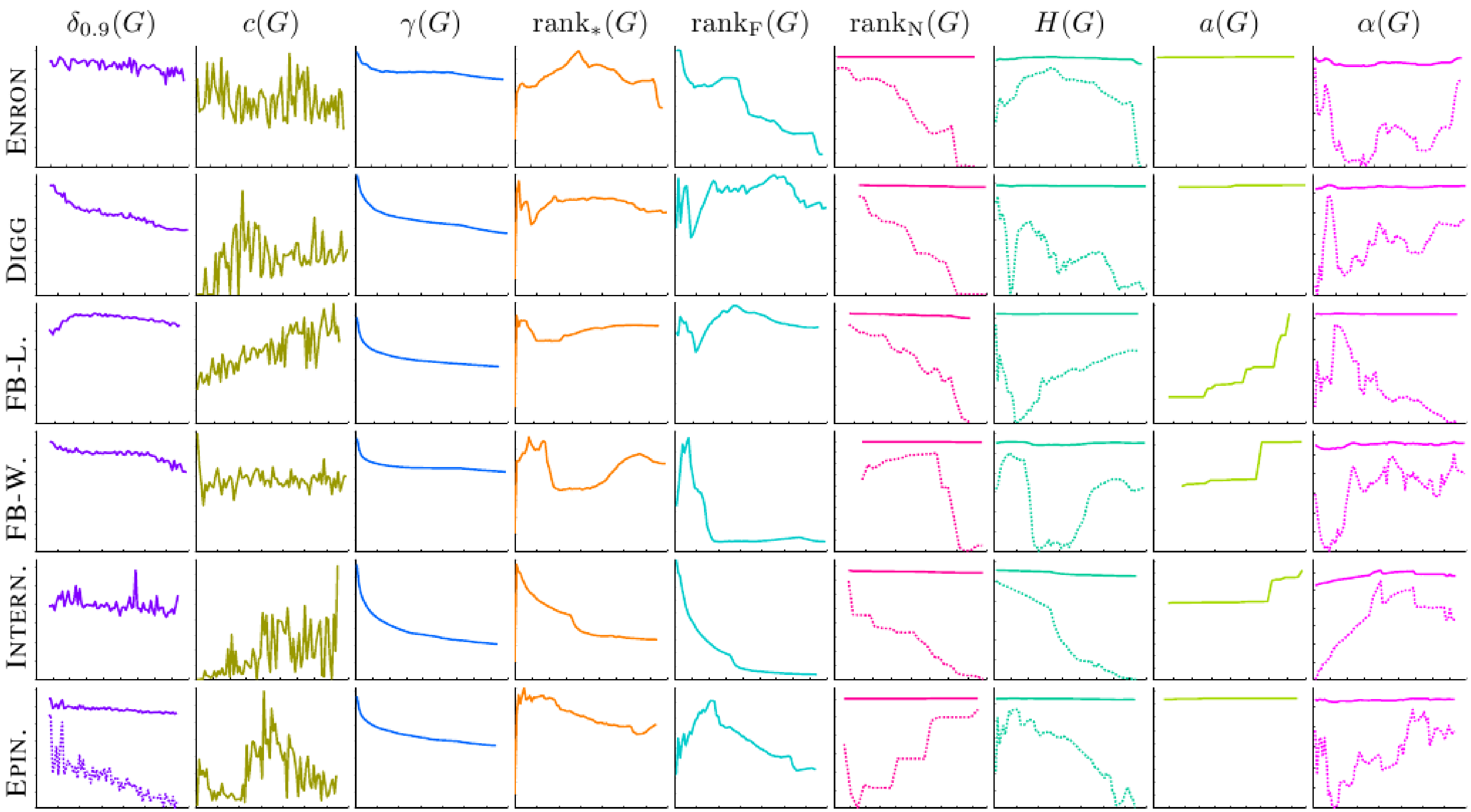
- Power-law exponent shrinks
  - Connection diversity shrinking
- Weighted spectral distribution shifts to zero
  - Emerging main components
- Entropy is constant
  - Effective number of communities is constant
- Network rank increases, then shrinks
  - Two-phase- model of expansion

- Dissemination: submitted to ICDM 2011

WeST People and Knowledge Networks

More Datasets

ROBUST

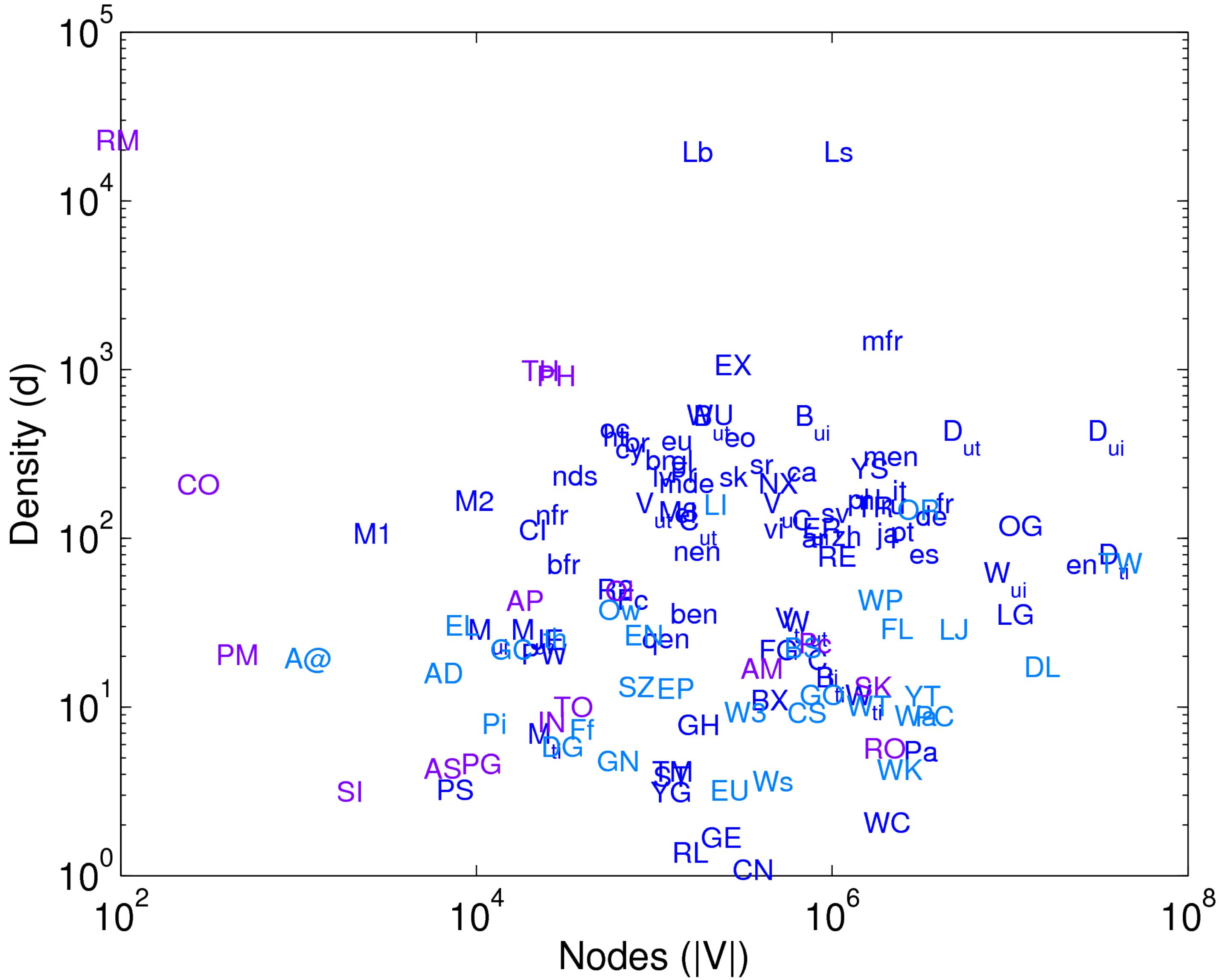


WeST People and Knowledge Networks

KONECT – Koblenz Network Collection

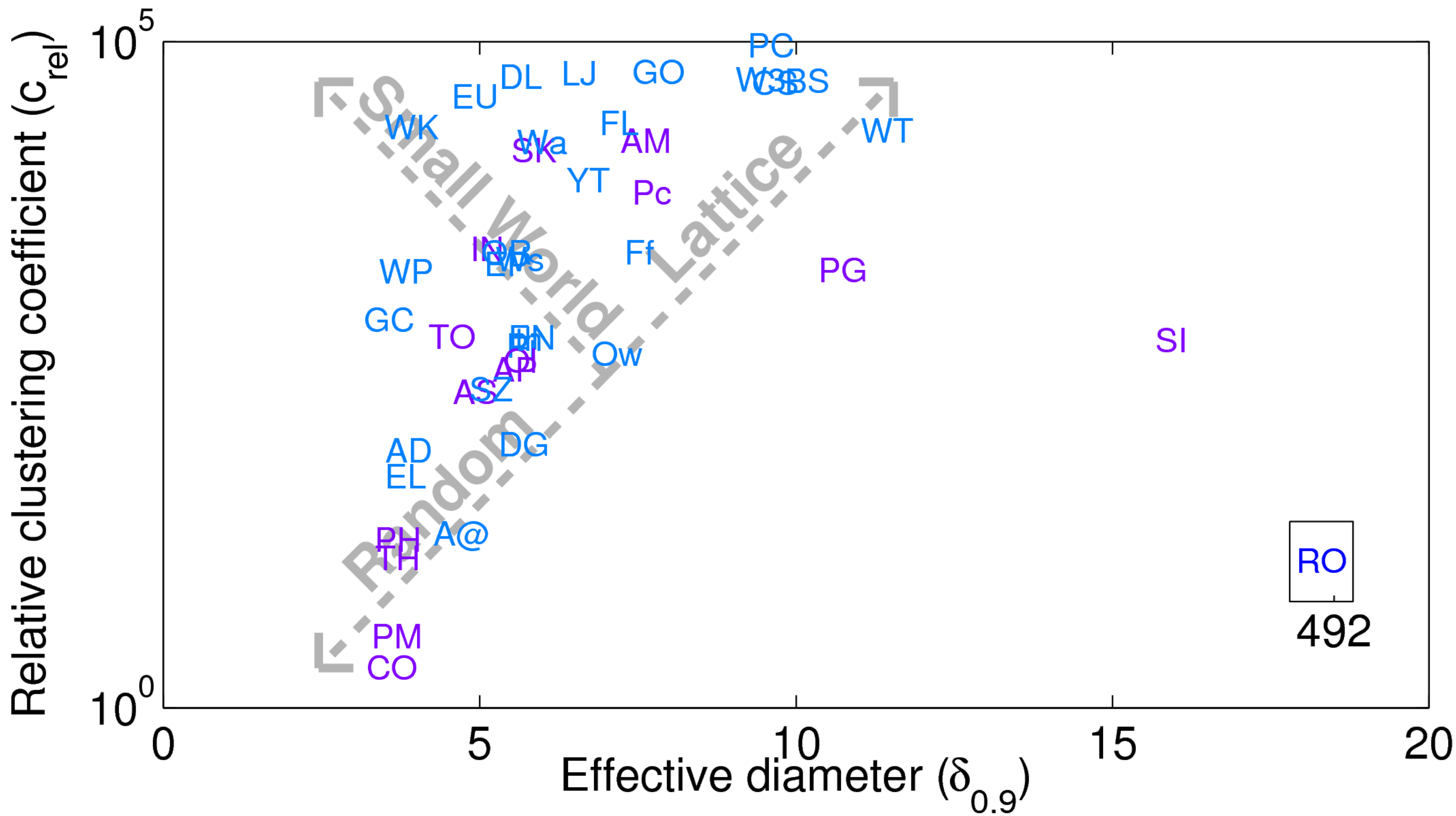
- 121 network datasets
- Categories: authorship, communication, co-occurrence, features, folksonomy, interaction, physical, ratings, reference, social, trust
- All code free/open

KONECT – Overview



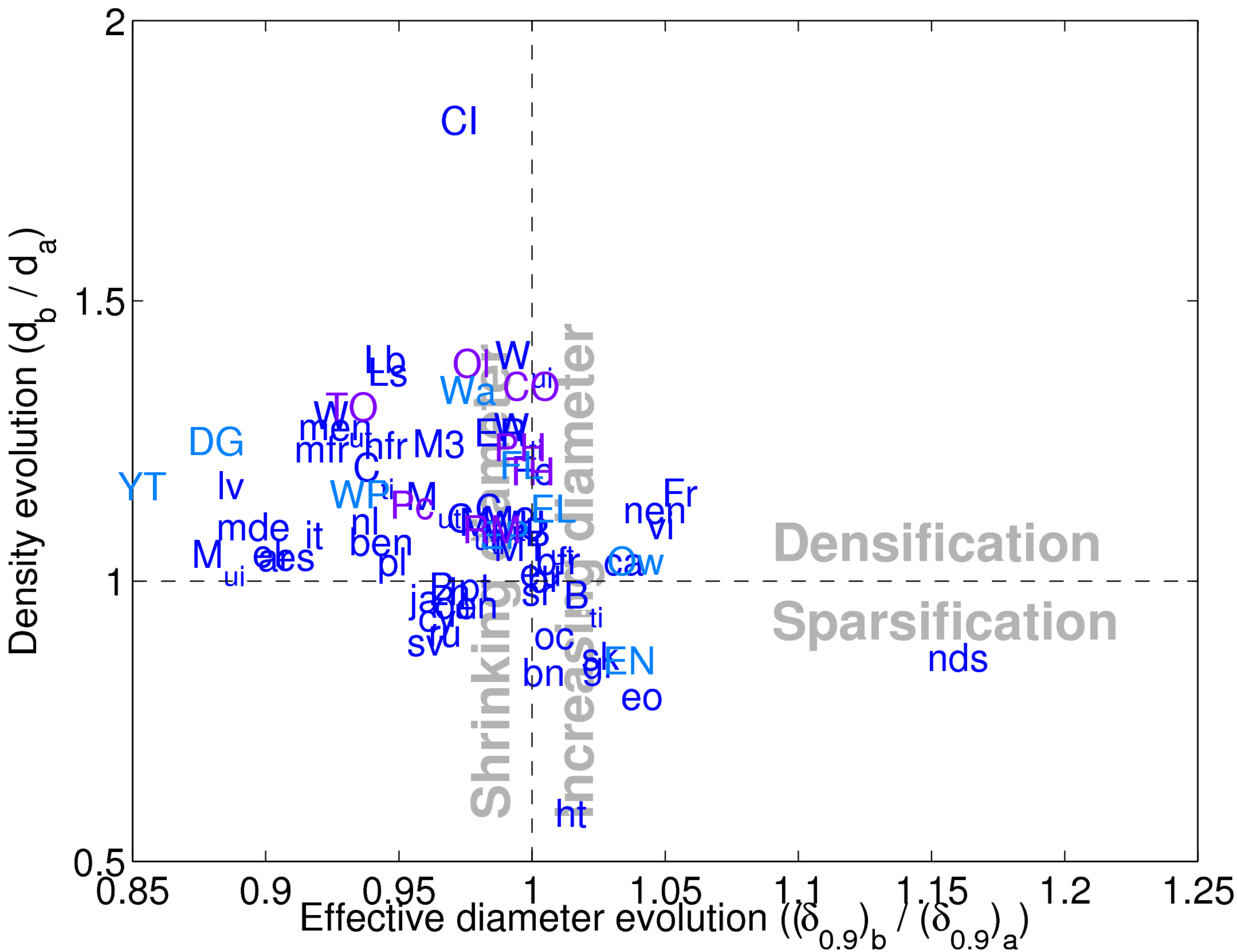
KONECT – Small-world Network Analysis

ROBUST



WeST
People and Knowledge Networks

KONECT – Dynamic Network Analysis



KONECT – Roadmap

- Release extraction code
- Website (including statistics, plots)
- Datasets (where legally possible)
- Analysis code (Matlab/Octave)

- Which datasets would you like to have?
- What datasets can you contribute?

References

Emamy K., Cameron R., 2007; CiteULike: A researcher's social bookmarking service; Ariadne, no. 51

Fay D., Haddadi H., Thomason A., Moore A. W., Mortier R., Jamakovic A., Uhlig S., Rio M., 2010; Weighted spectral distribution for Internet topology analysis: Theory and applications; IEEE Trans. Networking, vol. 18, no. 1, pp. 164–176

Klimt B., Yang Y., 2004; The Enron corpus: A new dataset for email classification research; Proc. European Conf. on Machine Learning, pp. 217–226

Kunegis J., Sizov S., 2011; Diversity vs uniformity: understanding the evolution of large networks; unpublished

Massa P., Avesani P., 2005; Controversial users demand local trust metrics: an experimental study on epinions.com community; Proc. American Association for Artificial Intelligence Conf., pp. 121–126

Newman M. E. J., 2006; Power laws, Pareto distributions and Zipf's law; Contemporary Phys., vol. 46, no. 5, pp. 323–351

Random Edge Order