SPATIAL NETWORK SURROGATES FOR DISENTANGLING COMPLEX SYSTEM STRUCTURE FROM SPATIAL EMBEDDING OF NODES

Marc Wiedermann, Jonathan F. Donges, Reik V. Donner and Jürgen Kurths

Motivation

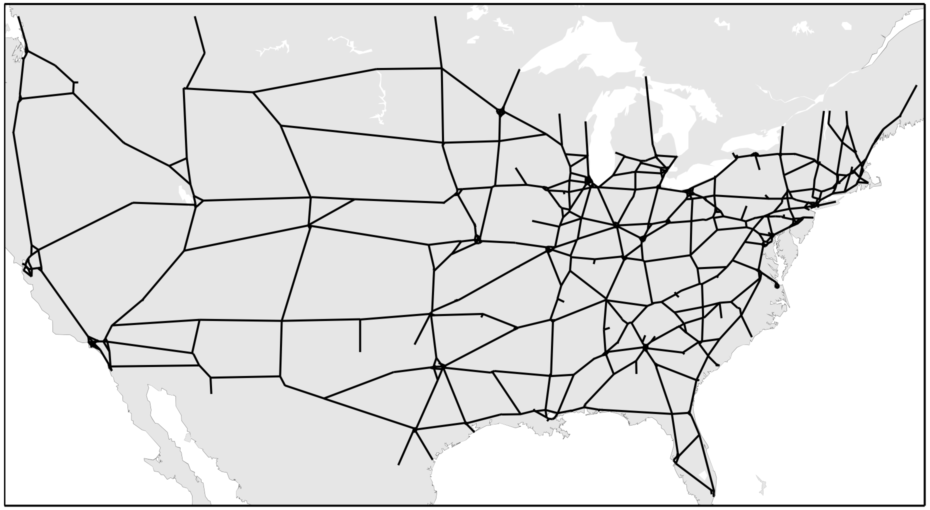
Many real-world complex networks are spatially embedded

- Transportation
- Communication
- Trade

Assessment of global network measures for:

- Characterization
- Intercomparison
- Classification (small-world networks)

US Interstate Network



Global Clustering coefficient = 0.1
Average Path length = 20

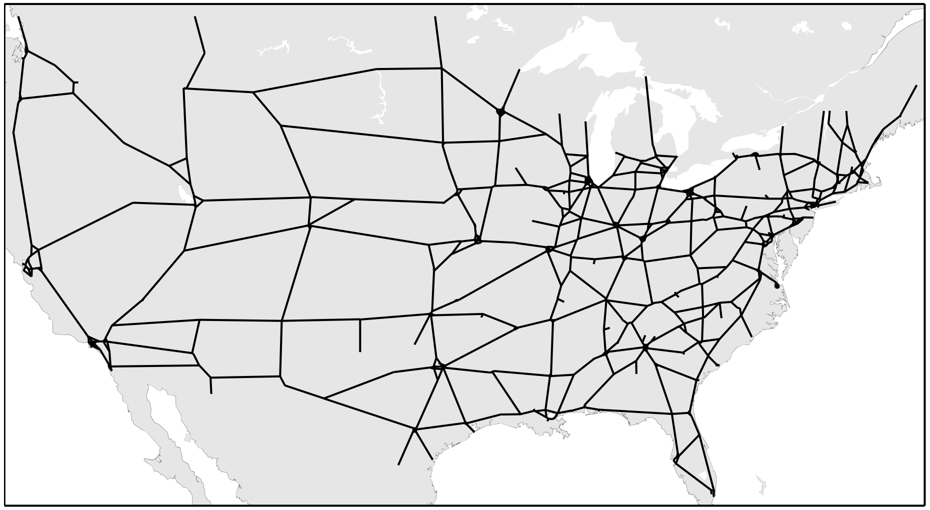
Motivation – Research questions

To what extent are global network characteristics predetermined by spatial embedding?

Given fundamental information on the network's spatial embedding can one estimate the average path length and global clustering coefficient?

Are there classes of networks that are affected by spatial embedding and others that are not?

US Interstate Network



Global Clustering coefficient = 0.1
Average Path length = 20

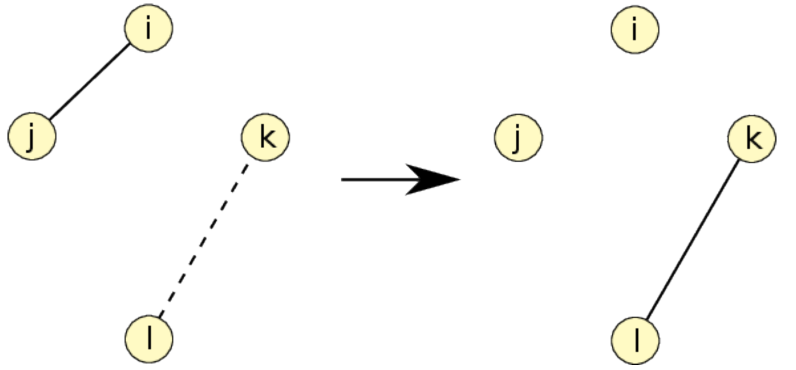
Plan of action

Create ensemble of surrogates by iteratively rewiring a given network and preserving "lower-level" topological and geographical features:

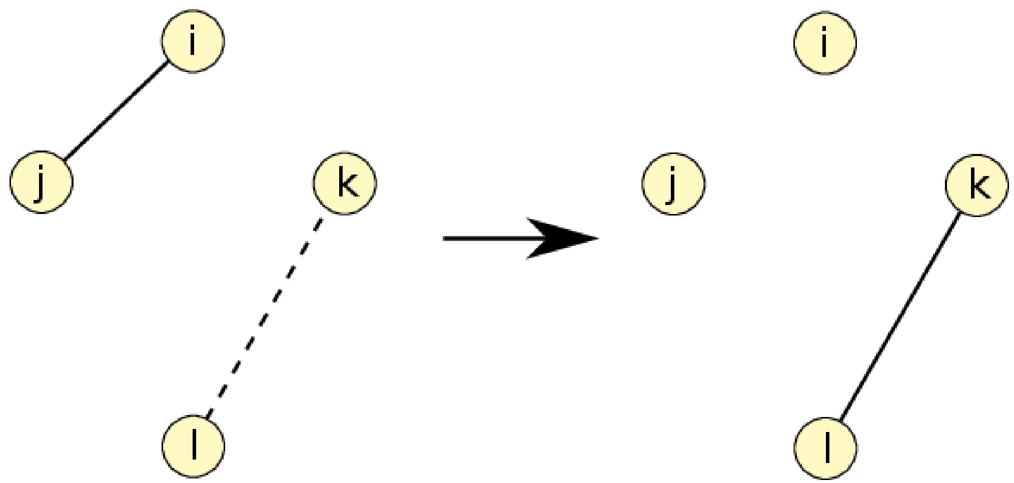
- Average degree K (as does the ER model)
- Local degree k_v (as does approximately the configuration model)

Introduce two new models that explicitly take into account spatial embedding:

- Link length distribution p(l)
  → GeoModel I
- Local link length distribution $p_v(l)$
  → GeoModel II



Random rewiring



$$\mathbf{C} = A'_{ij} \wedge \neg A'_{kl}$$

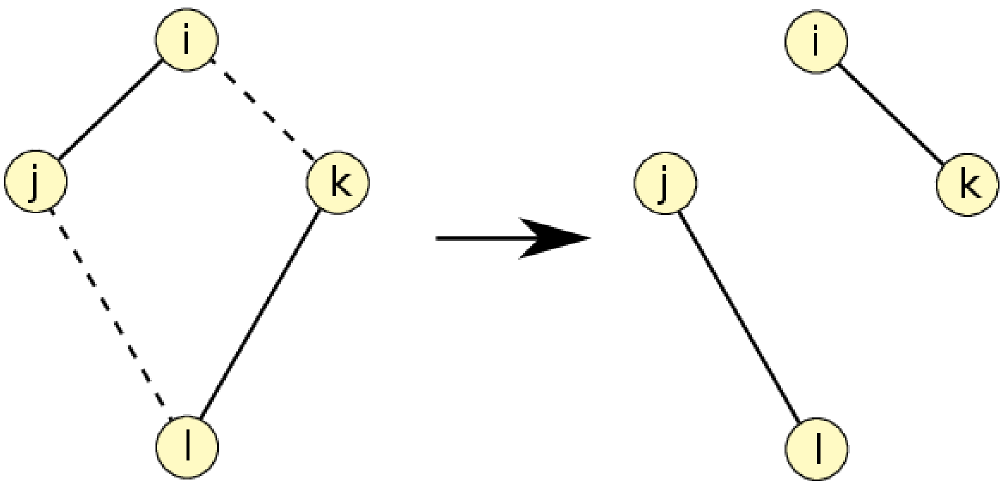
$$\mathbf{C} = \mathbf{C}_1$$

Preserves:

- Average degree K

Converges into ER random graph for sufficiently many steps

Random link switching



Preserves:

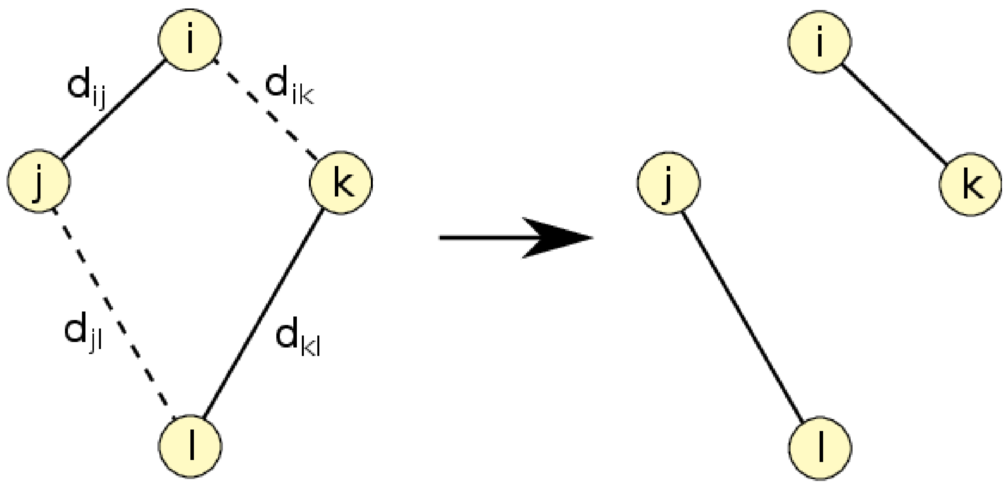
- Average degree K
- Local degree k_v

$$\mathbf{C}_1 = A'_{ij} \wedge A'_{kl} \wedge \neg A'_{ik} \wedge \neg A'_{jl}$$

Approximately converges into configuration model for sufficiently many steps

$$\mathbf{C} = \mathbf{C}_1$$

GeoModel I



Preserves:

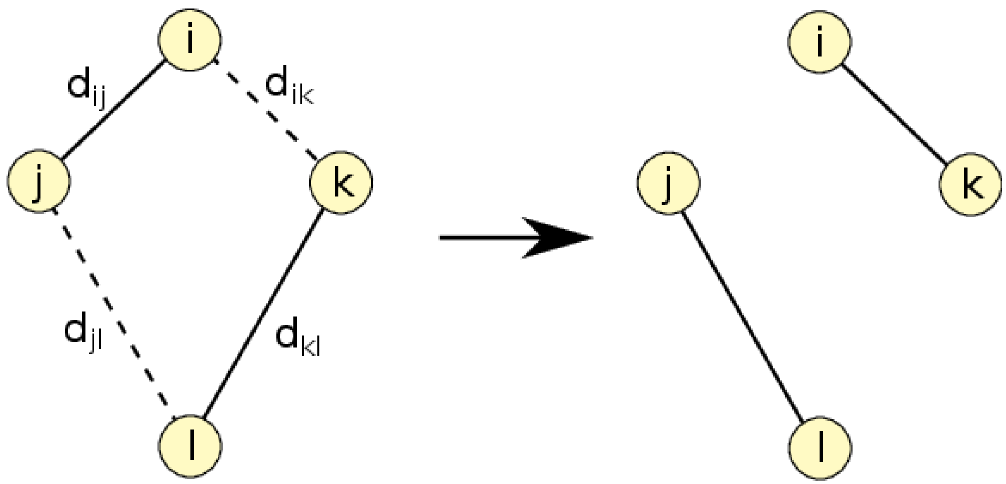
- Average degree K
- Local degree k_v
- Link length distribution p(l)

$$\mathbf{C}_1 = A'_{ij} \wedge A'_{kl} \wedge \neg A'_{ik} \wedge \neg A'_{jl}$$

$$\mathbf{C}_2 = \Theta(\epsilon d_{ij} - |d_{ij} - d_{ik}|) \wedge \Theta(\epsilon d_{kl} - |d_{kl} - d_{jl}|)$$

$$\mathbf{C} = \mathbf{C}_1 \wedge \mathbf{C}_2$$

GeoModel I



Preserves:

- Average degree K
- Local degree k_v
- Link length distribution p(l)

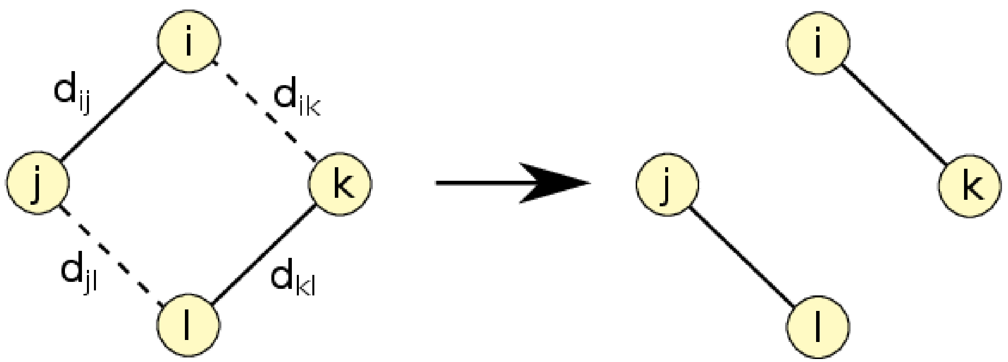
$$\mathbf{C}_1 = A'_{ij} \wedge A'_{kl} \wedge \neg A'_{ik} \wedge \neg A'_{jl}$$

$$\mathbf{C}_2 = \Theta(\epsilon d_{ij} - |d_{ij} - d_{ik}|) \wedge \Theta(\epsilon d_{kl} - |d_{kl} - d_{jl}|)$$

$$\mathbf{C} = \mathbf{C}_1 \wedge \mathbf{C}_2$$

Tolerance (only free parameter of the model)

GeoModel II



Preserves:

- Average degree K
- Local degree k_v
- Link length distribution p(l)
- Local link length distribution $p_v(l)$

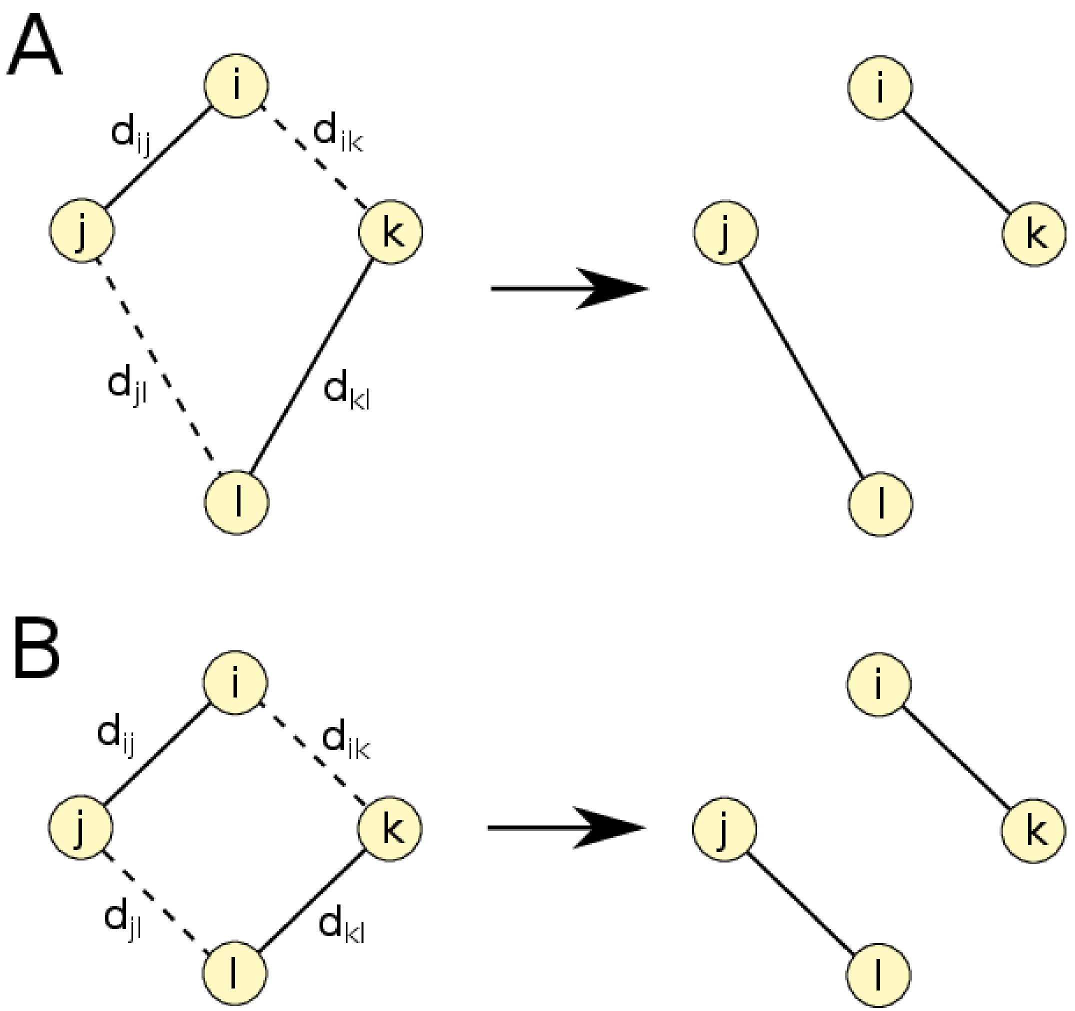
$$\mathbf{C}_1 = A'_{ij} \wedge A'_{kl} \wedge \neg A'_{ik} \wedge \neg A'_{jl}$$

$$\mathbf{C}_2 = \Theta(\epsilon d_{ij} - |d_{ij} - d_{ik}|) \wedge \Theta(\epsilon d_{kl} - |d_{kl} - d_{jl}|)$$

$$\mathbf{C}_3 = \Theta(\epsilon \max(d_{ik}, d_{jl}) - |d_{ik} - d_{jl}|)$$

$$\mathbf{C} = \mathbf{C}_1 \wedge \mathbf{C}_2 \wedge \mathbf{C}_3$$

Rewiring scheme - Summary



Search for certain geometries in the networks:

- Rewiring along kites preserves global link length distribution
- Rewiring along diamonds preserves local link length distribution

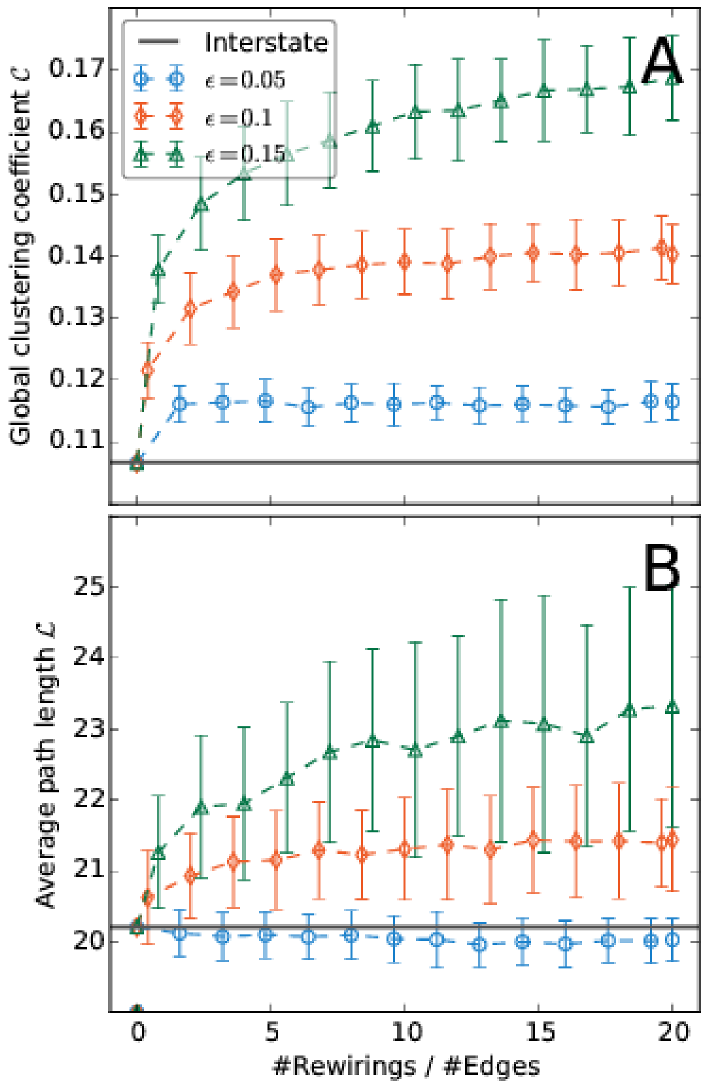
Both models exhibit only one free parameter (the tolerance)

Data

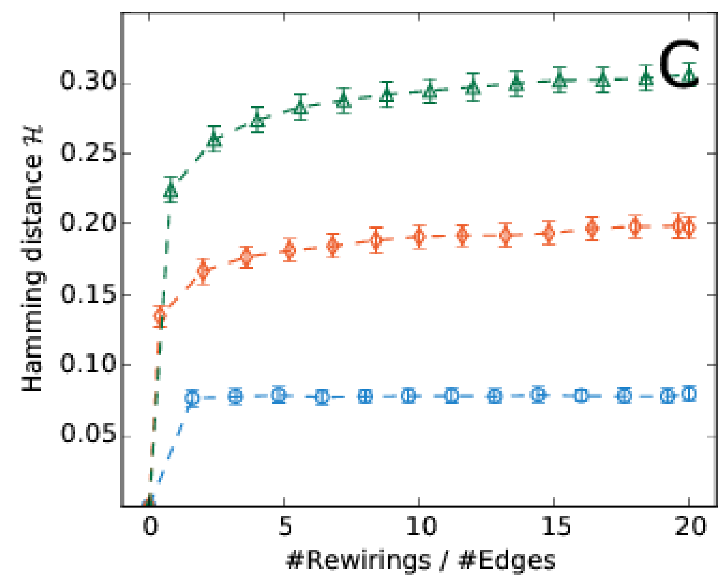
- 6 real world networks
- 2 random networks as benchmark

Name	N	M	K	ρ	$\mathcal{C}$	$\mathcal{L}$
US airline	190	837	8.86	0.0466	0.679	2.176
Internet	13.372	28.253	4.23	0.0003	0.423	3.630
US interstate	935	1.315	2.82	0.0030	0.107	20.207
Scandinavian power grid	236	318	2.71	0.0115	0.084	9.156
World trade	186	7.043	76.14	0.4094	0.815	1.594
Urban roads (Eschwege)	855	1.174	2.75	0.0032	0.050	18.313
Random geometric graph	2.000	5.493	5.50	0.0027	0.588	30.428
Erdős-Rényi graph	2.000	5.493	5.50	0.0027	0.003	4.643

Influence of tolerance

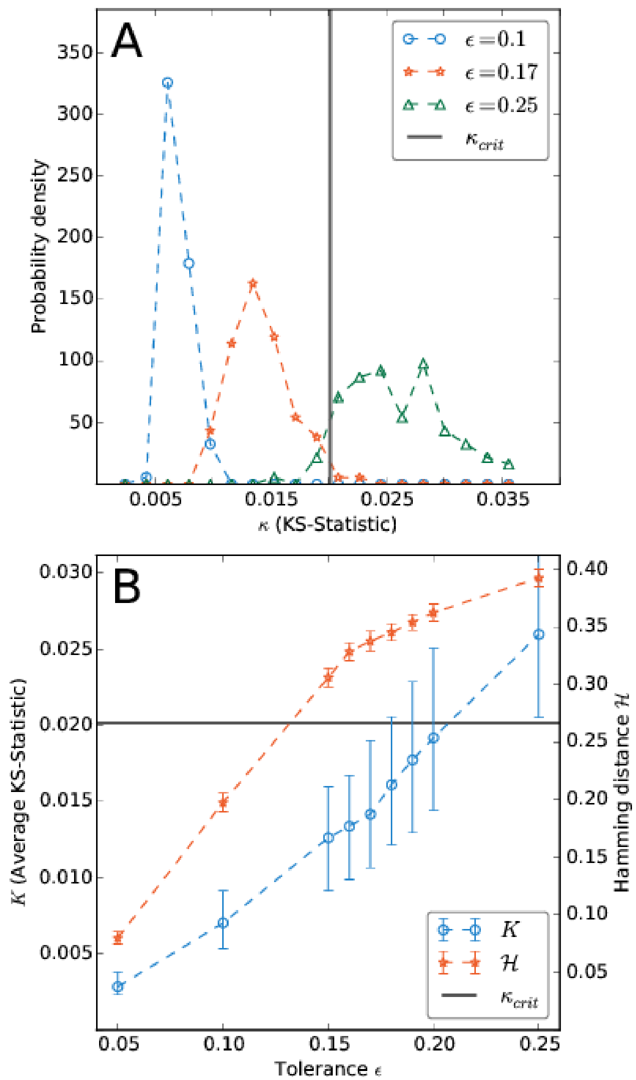


- Surrogates' global characteristics deviate from target value with increasing tolerance
- Find tolerance such that hamming distance is maximized but link length distributions are still statistically indistinguishable



$$\mathcal{H} = \frac{1}{4M} \sum_{i,j} |A'_{ij} - A_{ij}| \in [0, 1]$$

Estimating the tolerance

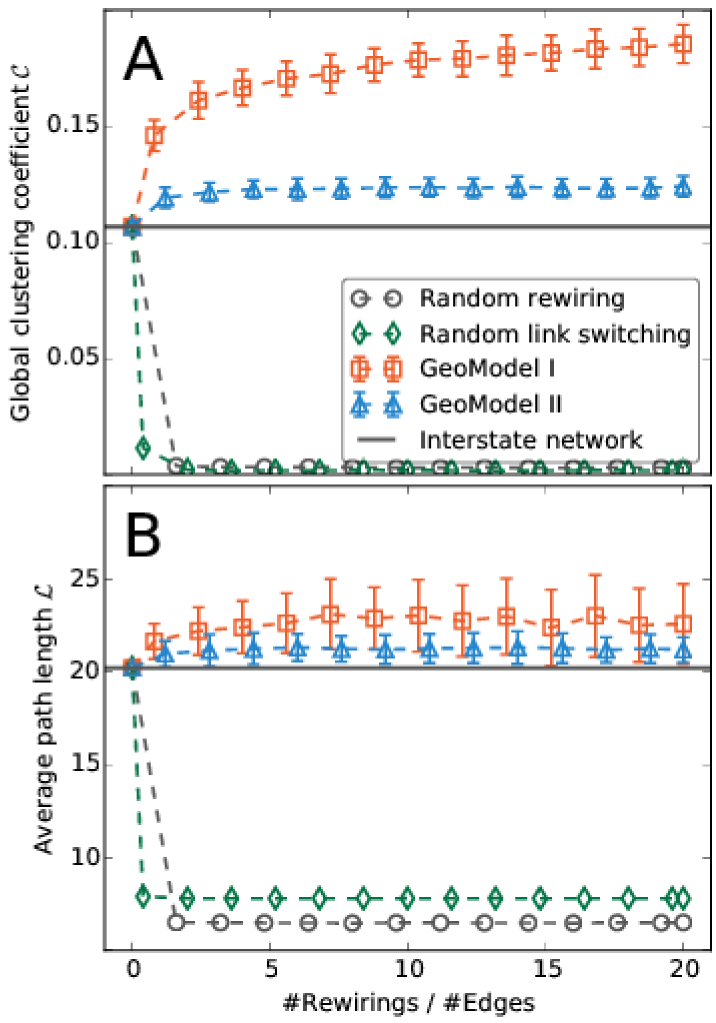


Create an ensemble of surrogate networks for each choice of tolerance

Assess distribution of KS-statistics

Choose tolerance such that 95% of surrogates' link length distributions are statistically indistinguishable from original value at 95% confidence

Results – Interstate network



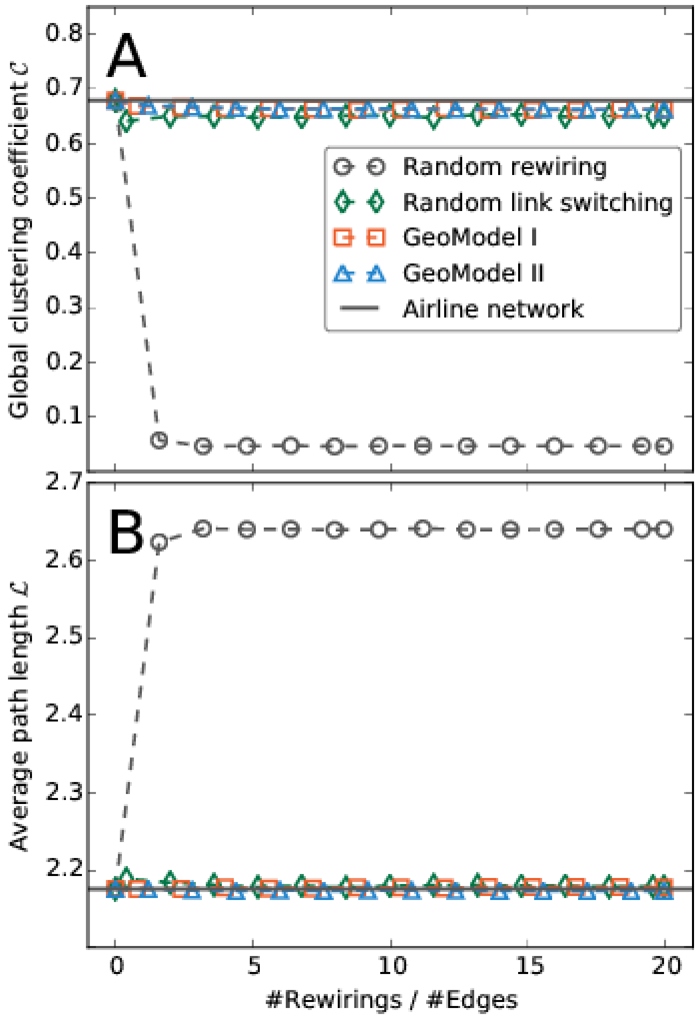
Average path length and global clustering of surrogates created from random rewiring/link switching **deviate** from that of original network

Taking into account spatial embedding:

GeoModel I reproduces well the average path length

GeoModel II also recaptures well the global clustering coefficient

Results – Airline network



<u>In contrast to interstate network:</u>

Random link switching already reproduces well global characteristics of original network

→ Spatial embedding of nodes needs not necessarily taken to be into account

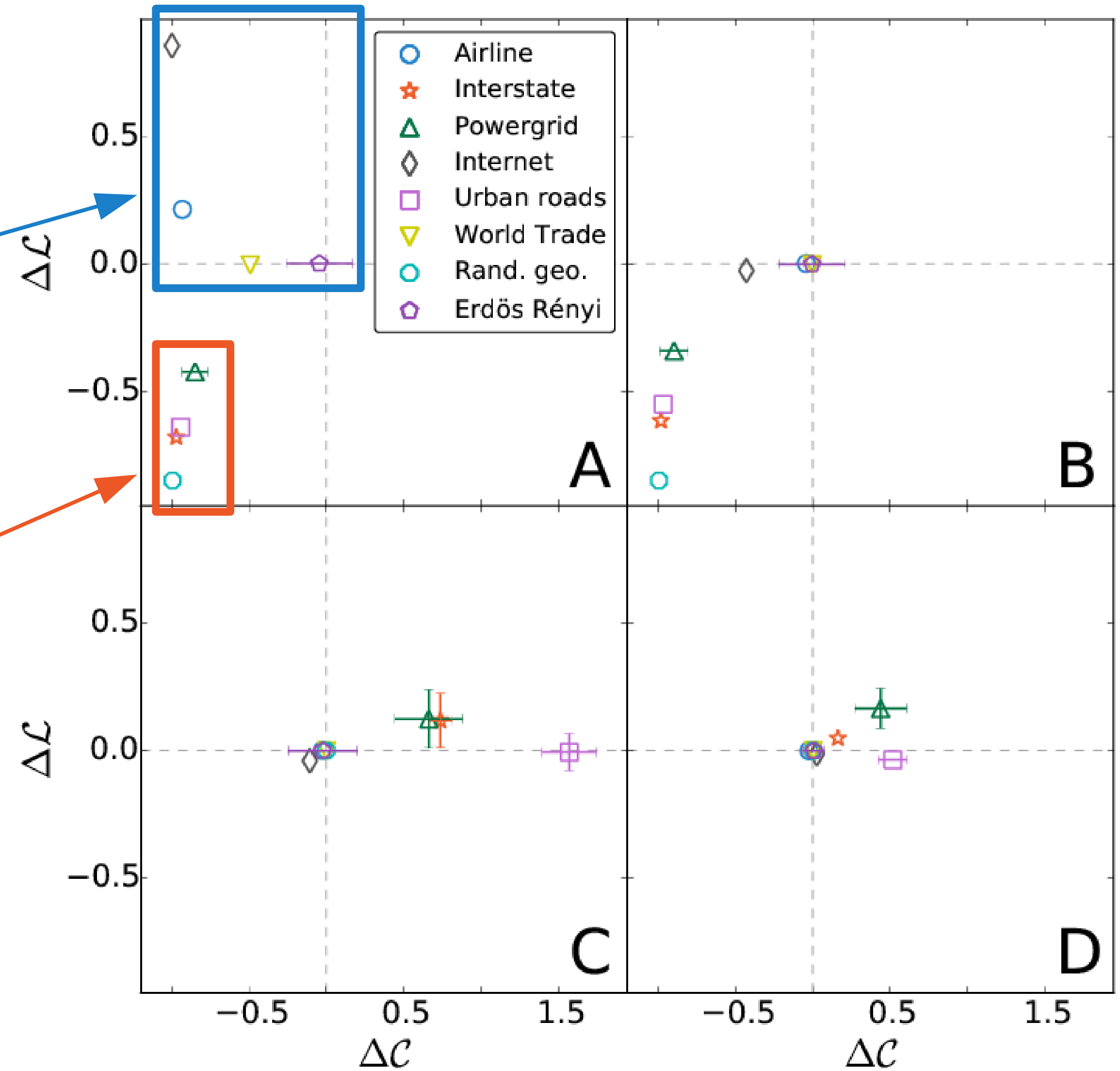
Are there different classes of spatially embedded networks?

Intermediate Summary

No spatial network under study is purely random

Random link switching (preserving local degree) recaptures average path length and global clustering for certain networks

For a second class only GeoModel I and II (taking into account spatial embedding) recapture original networks' global characteristics



Link length distributions

Is it possible to get a similar discrimination from assessing the distributions of link lengths only?

We can measure:

- probability for a distance between nodes given a link
- probability for a distance between nodes

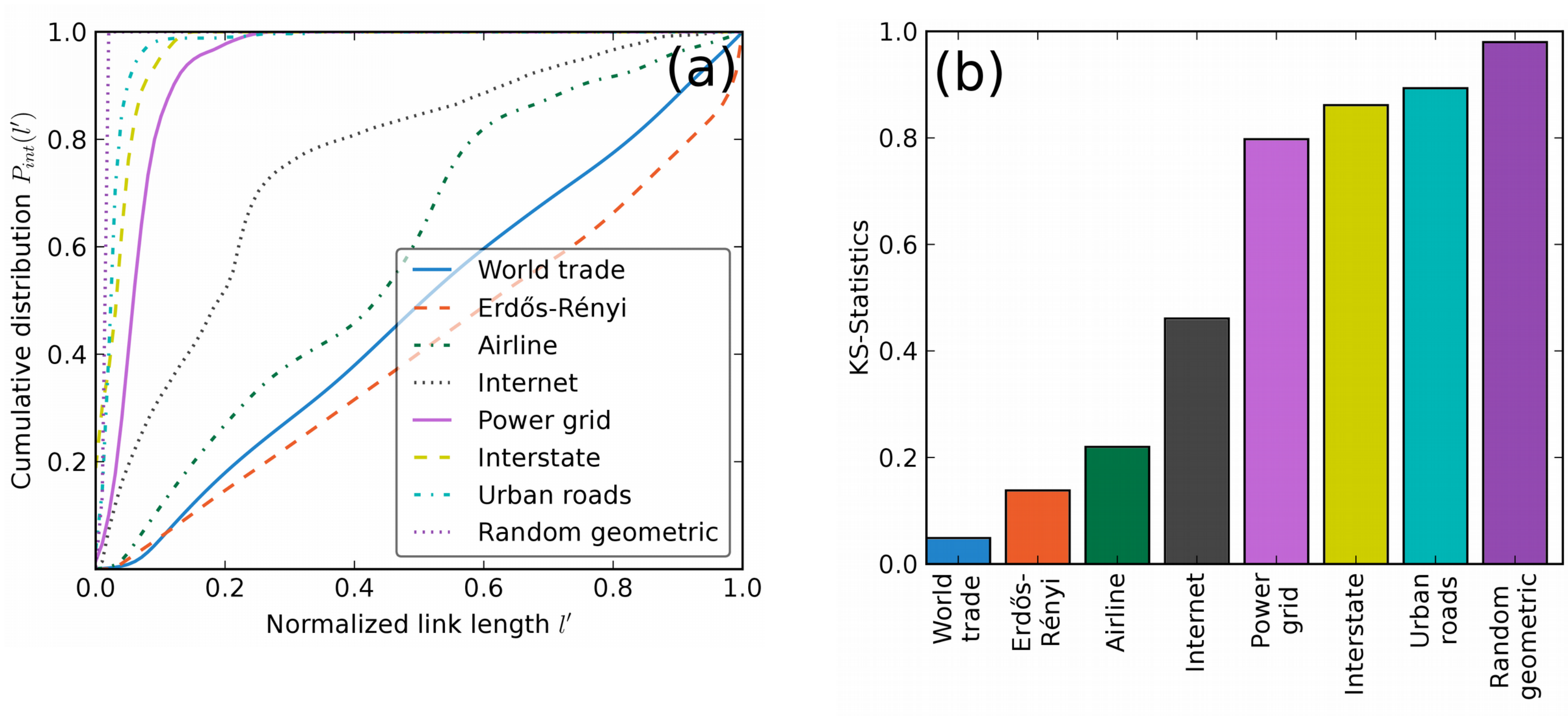
We are interested in the probability of the distance between nodes given a link and call this the intrinsic linking probability

$$p(A_{ij}=1|d_{ij})p(d_{ij}) = p(d_{ij}|A_{ij}=1)p(A_{ij}=1)$$

$$\Rightarrow p(A_{ij}=1|d_{ij}) \propto \frac{p(d_{ij}|A_{ij}=1)}{p(d_{ij})}$$

$$p(\text{link is present} \mid \text{distance}) \propto \frac{p(\text{distance} \mid \text{link is present})}{p(\text{distance})}$$

Link length distributions

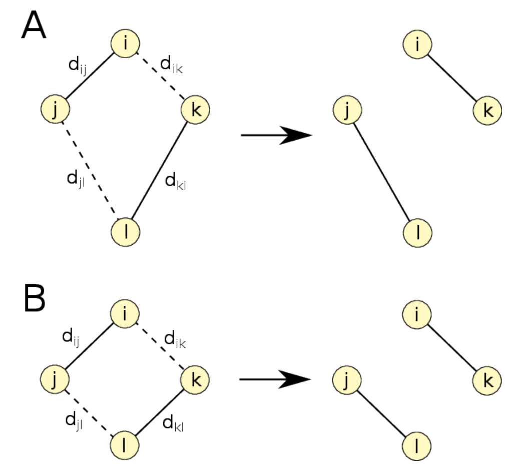


- Intrinsic linking probabilities of the world trade, Erdős–Rényi and airline networks are close to random uniform
- Intrinsic linking probabilities of the Random geomatric graph, the power grid, interstate, and urban road networks are close to exponential

Confirms the classifications from the GeoModels

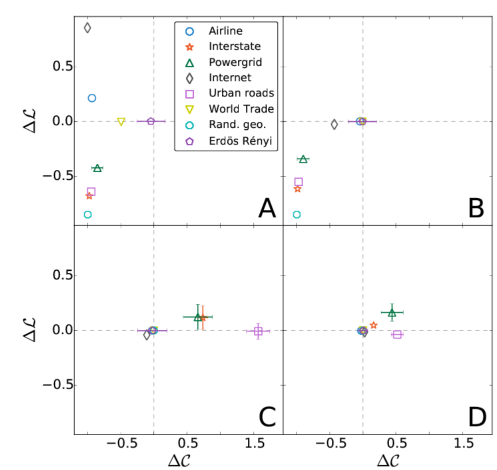
Conclusion

Introduced two random network models that take into account spatial embedding of nodes



Identified set of networks for which spatial embedding should be taken into account when assessing global network characteristics

In most cases global measures are to high degree explainable by spatial embedding → Small-World effect may be explainable in the same way in many networks (Bialonski et al., Chaos, 2010)



Reference

PHYSICAL REVIEW E **93**, 042308 (2016)

Spatial network surrogates for disentangling complex system structure from spatial embedding of nodes

Marc Wiedermann,[1,2,*] Jonathan F. Donges,[1,3] Jürgen Kurths,[1,2,4,5] and Reik V. Donner[1]

[1]*Potsdam Institute for Climate Impact Research, P.O. Box 60 12 03, 14412 Potsdam, Germany, EU*

[2]*Department of Physics, Humboldt University, Newtonstraße 15, 12489 Berlin, Germany, EU*

[3]*Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, 114 19 Stockholm, Sweden, EU*

[4]*Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3FX, United Kingdom, EU*

[5]*Department of Control Theory, Nizhny Novgorod State University, Gagarin Avenue 23, 606950 Nizhny Novgorod, Russia*

(Received 26 September 2015; revised manuscript received 22 January 2016; published 12 April 2016)

Physical Review E 93, 042308 (2016)

doi: 10.1103/PhysRevE.93.042308