Complex Networks and Hidden Metric Spaces

M. ÁNGELES SERRANO
Dpt. Física Fonamental
Universitat de Barcelona

Funding
Mapping complexity

Through history, **maps** have been at the center of political, economic and geostrategic decisions to become a critical piece in our every day lives.

- A way of storing and presenting information and communicating findings.
- Integral information source, visually appealing.
- Locational distributions and spatial relationships
- Conceptualization of patterns and processes that operate through space.

Our **goal** is to **map complex real systems** in an **embedding metric space** that ought not to be geographical or spatially obvious but that may be a condensate of different intrinsic attributes determining how distant/similar the elements of the system are.
Self-similarity

Scale invariance

Exact form of self-similarity

\[ f(\lambda v) = \lambda^s f(v) \]
\[ P(\lambda k) = \lambda^{-\gamma} P(k) \]

Self-similarity

The same properties at different length scales (exact or approximate or statistical)

Fractality

Self-similarity is a property of fractals, objects with a length that depends on the measuring scale (recurrent detailed patterns)
Self-similarity - Fractality

Self-similarity is a property of fractals. **Fractals**: Objects with Hausdorff dimension greater than its topological dimension (usually non-integral). The measured length depends on the measuring scale.

In fractal geometry, box-counting is the primary way to evaluate the fractal dimension of a fractal.
Self-similarity - Fractality

Topological self-similarity
Box-covering renormalization in complex networks

The network is tiled with boxes such as all nodes in a box are connected by a minimum distance smaller than a given $\ell_B$.

Each box is replaced by a single node and two renormalized nodes are connected if there is at least one link between the original boxes.

Self-similarity - Fractality

Topological self-similarity
Box-covering renormalization in complex networks

Some systems (WWW, biological networks) are \textit{fractal} and have degree distributions that remain invariant and have finite “fractal” dimension

\[ N_B(\ell_B)/N \sim \ell_B^{-d_B} \quad \text{and} \quad k_B(\ell_B)/k_{hub} \sim \ell_B^{-d_k} \]

Some systems (the Internet) are \textit{non-fractal}, with scaling laws that are replaced by exponentials

\[ d_B \to \infty \quad \text{and} \quad d_k \to \infty \]

\[ N_B(l_B)/N \sim \exp\left(-\left(\ln n/2\right)\ell_B\right) \]

\[ k_B(l_B)/k_{hub} \sim \exp\left(-\left(\ln s/2\right)\ell_B\right) \]

Self-similarity - Fractality

Specific PROBLEMS of the box-covering methodology
• The box covering partition is not univocal, results may depend on the specific partition of nodes into boxes
• Large fractal dimension may not be distinguishable from exponential behavior
• Small range in which the scaling is valid small world property
• Self-similarity of the degree distribution under renormalization, but in general correlations do not scale

Other proposals
• J S Kim, K-I Goh, B Kahng and D Kim, Fractality and self-similarity in scale-free networks, New J. Phys. 9 177 (2007) a slightly different box covering fractality and self-similarity are disparate notions in SF networks the Internet is a non-fractal SF network, yet it exhibits self-similarity
• J I Alvarez-Hamelin, L Dall'Asta, A Barrat, and A Vespignani, K-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases, Networks and Heterogeneous Media, 3(2):371-293 (2008) the Internet shows statistical self-similarity of the topological properties of k-cores
Self-similarity of complex networks

Self-similarity as the scale-invariance of the degree distribution

Self-similarity of complex networks is not well defined yet in a proper geometric sense

Lack of a metric structure except lengths of shortest paths, too homogeneous

Small world property $N \approx e^{l/l_0}$

geometric length scale?
Underlying metric spaces

Geography as an obvious geometrical embedding: airport networks, urban networks…

Hidden metric spaces:
WWW (similarity between pages induced by content), social networks (closeness in social space)…

…how to identify hidden metric spaces and their meaning…
Underlying metric spaces

We conjecture that hidden geometries underlying some real networks are a plausible explanation for their observed self-similar topologies.

- Some real scale-free networks are self-similar (degree distribution, degree-degree correlations, and clustering) with respect to a simple degree-thresholding renormalization procedure (purely topological).

- A class of hidden variable models with underlying metric spaces are able to accurately reproduce the observed topology and self-similarity properties.

Self-similarity of real complex networks

degree-thresholding renormalization procedure \( k > k_T \)

\[ k_i / < k_i (k_T) > \]

\( \frac{3}{2.6} \) \( \frac{2}{2.7} \) \( \frac{2}{2.3} \) \( \frac{0}{1.3} \)

\( k_T = 0 \) \( k_T = 1 \) \( k_T = 2 \) \( k_T = 3 \)

G \hspace{2cm} G(k_T)
## Self-similarity of real complex networks

### BGP map of the Internet at the AS level
- SF with exponent 2.2
- $N=17446$
- $<k>=4.68$

### PGP social web of trust
- SF with exponent 2.5
- $N=57243$
- $<k>=2.16$

(also U.S. airports network, but not so challenging since it is embedded in geo)

A random model like the CM will produce self-similar networks regarding the degree distribution and degree-degree correlations, if the degree distribution of the complete graph is SF...
Self-similarity of real complex networks

Degree-dependent clustering

FIG. 2 (color online). (a)–(d) Degree-dependent clustering coefficient as a function of the rescaled internal degree for the Internet BGP map, the PGP web of trust, and their randomized versions. (e) Average clustering coefficient as a function of the threshold degree $k_T$ for renormalized real networks and their randomized counterparts. (f) Internal average degree as a function of $k_T$ for the same networks.
Networks meet geometry

**CLUSTERING** is the first property of networks non-trivially self-similar.

The key point to model self-similar networks is to reproduce self-similar clustering.

The clue for the connection between complex networks and metric spaces is clustering.

Why is clustering important?

\[ AC \leq AB + BC \]

**TRIANGLE INEQUALITY**

for any triangle, the sum of the lengths of any two sides must be greater than the length of the remaining side.

---

The connection probability is an integrable function of the effective distance, which depends on the expected degrees.

\[ d_c(\kappa, \kappa') \propto (\kappa \kappa')^{1/D} \]

Nodes that are close to each other are more likely to be connected. Nodes with larger degrees can reach further.

**Networks meet geometry**

class of hidden variable models with underlying metric spaces

The connection probability is an integrable function of the effective distance, which depends on the expected degrees.

\[ p(\kappa) \]

\[ P \left( \frac{d}{d_c} \right) \]

**Connectivity law \( \rightarrow \) “Gravity law”**

The S1 model

\[ p(\kappa) = \frac{1}{1 + (d_{mr}/\mu \kappa m \kappa r)^\beta} \]

\[ d(\theta, \theta') = \Delta \theta \frac{N}{2\pi} \]

\[ N = \delta 2\pi R \]

\[ k \rightarrow \text{popularity} \]

\[ d \rightarrow \text{similarity} \]
The $S_1$ model - specifications

- $\rho(\kappa)$ controls the degree distribution, SF $\rho(\kappa) \approx P(k)$
- independently, $\beta$ controls the level of clustering, strong clustering
- given $\beta$, parameter $\mu = \left(\frac{\gamma-2}{\gamma-1}\right)^2 \frac{(\alpha-1) < k >}{2\delta \kappa_0^2}$ controls the average degree
- if $1 < \beta < 2$ or $2 < \gamma < 3$, small-world!!! but underlying metric!!!
Self-similarity of the S1 model

Thresholding of the hidden variable $\kappa$

\[
\kappa_0 \rightarrow \kappa_T; \quad \delta \rightarrow \delta \left[ \frac{\kappa_T}{\kappa_0} \right]^{1-\gamma}
\]
Self-similarity of the $S_1$ model

The ensemble is self-similar with respect to $T$ if any subgraph belongs to the original ensemble.

Thresholding of the hidden variable $\kappa$

$$G_T(\{\alpha\}) = G(\{\alpha_T\})$$
Self-similarity of the S1 model

\[ \rho(\kappa) = (\gamma - 1) \frac{K_0^{\gamma-1}}{K^\gamma} \]

\[ N = \delta 2\pi R \]

\[ r(\theta, \kappa; \theta', \kappa') = \left(1 + \frac{d(\theta, \theta')}{\mu \kappa \kappa'}\right)^{-\alpha} \]

\[ \mu = \left(\frac{\gamma - 2}{\gamma - 1}\right) \frac{(\alpha - 1) \langle k \rangle}{2\delta \kappa_0^2} \]

\[ \rho_r(\kappa) = (\gamma - 1) \frac{K_T^{\gamma-1}}{K^\gamma} \]

\[ N_r = \delta_r 2\pi R \]

\[ N_r = N \int_{K_r}^\infty \rho(\kappa) d\kappa = N \left(\frac{K_0}{K_T}\right)^{\gamma-1} \]

\[ r(\theta, \kappa; \theta', \kappa') = \left(1 + \frac{d(\theta, \theta')}{\mu \kappa \kappa'}\right)^{-\alpha} \]

\[ \mu = \left(\frac{\gamma - 2}{\gamma - 1}\right) \frac{(\alpha - 1) \langle k \rangle_r}{2\delta_r \kappa_T^2} \]

\[ \mu = \left(\frac{\gamma - 2}{\gamma - 1}\right) \frac{(\alpha - 1) \langle k \rangle_r}{2\delta \kappa_0^2} \]

\[ \langle k \rangle_r = \langle k \rangle \left(\frac{K_T}{K_0}\right)^{3-\gamma} \]
In particular, clustering spectrum and clustering…

\[ c(\kappa | \kappa_T) = f(\kappa/\kappa_T), \quad \tilde{c}(k_i | \kappa_T) \approx f(k_i / \kappa_T^{3-\gamma}) = \frac{\tilde{f}[k_i / \langle k_i(k_T) \rangle]}{\langle k_i(k_T) \rangle}, \]

\[ \tilde{c}(\kappa_T) = \int_{\kappa_T} d\kappa \rho(\kappa | \kappa_T) \tilde{c}(\kappa | \kappa_T), \quad \text{Independent of } \kappa_T \]
Self-similarity of the S1 model
The $S_1$ model

- Topology generator
- Efficient communication without global knowledge
- Isomorphic to $H2$ in hyperbolic space
- Real networks can be mapped (embedded in geometry)
The S1 model – Topology generator

FIG. 1: Degree distribution $P(k)$, average nearest neighbours’ degree $k_{nn}(k)$, and degree-dependent clustering coefficient $\bar{c}(k)$ generated by our model with $\gamma = 2.1$ and $\alpha = 2$ compared to the same metrics for the real Internet map as seen by BGP data and the DIMES project.

FIG. 2: Degree distribution $P(k)$, average nearest neighbours’ degree $k_{nn}(k)$, and degree-dependent clustering coefficient $\bar{c}(k)$ generated by our model with $\gamma = 1.6$, $\alpha = 5$ and a cut-off at $k_c = 180$ compared to the same metrics for the real US airport network.
Navigability of complex networks

The underlying metric structure makes complex networks able to support efficient communication without global knowledge

Communication in terms of greedy routing

Greedy algorithm: finding the global optimum making the locally optimal choice at each stage

Greedy routing: navigation technique in which a node passes information to the neighboring node that is closest to the final destination

Boguna, M; and Krioukov, D., PRL 102, 058701 (2009)
Navigability of complex networks

Efficient communication means finding the shortest paths without global knowledge.

In navigable networks, greedy paths are shortest paths (in terms of topology).

**S1 networks** (and real networks) are **Navigable** because of the connection between link existence probability and hidden distance between nodes.

Explanation: zoom-out/zoom-in mechanism

Boguna, M; Krioukov, D; Claffy, KC
NATURE PHYSICS 5, 74-80 (2009)
S1 Isomorphic to H2 in hyperbolic space

S1 Newtonian (gravity law)

\[ d(\theta, \theta') \]
\[ d = \Delta \theta \frac{N}{2\pi} \]

Hyperbolic distance

\[ \cosh x_{ij} = \cosh r_i \cosh r_j - \sinh r_i \sinh r_j \cos \Delta \theta_{ij} \]

H2 Einstenian or relativistic (purely geometric)
## S1 Isomorphic to H2 in hyperbolic space

<table>
<thead>
<tr>
<th>Geometry</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperbolic geometry</td>
<td>Constant negative curvature</td>
</tr>
<tr>
<td></td>
<td>Every point is a saddle point</td>
</tr>
<tr>
<td></td>
<td>An infinite number of parallel lines</td>
</tr>
<tr>
<td>Euclidean geometry</td>
<td>Flat, curvature 0</td>
</tr>
<tr>
<td></td>
<td>Pairs of parallels are unique</td>
</tr>
<tr>
<td>Spherical geometry</td>
<td>Constant positive curvature</td>
</tr>
<tr>
<td></td>
<td>No parallel lines</td>
</tr>
</tbody>
</table>

Diagram showing
- Hyperbolic geometry
- Euclidean geometry
- Spherical geometry
Hyperbolic spaces are difficult to envisage because they cannot be isometrically embedded into any Euclidean space. The reason is that the former are “larger” and have more “space” than the latter.

The Poincaré tool by Bill Horn can be used to construct a tessellation of the hyperbolic plane in the Poincaré disc model given a tiling image, http://poincare.sourceforge.net/
**S1 Isomorphic to H2 in hyperbolic space**

**S1 Newtonian** (gravity law)

\[ p \left( \frac{d}{\mu k k'} \right) = p \left( \frac{\Delta \theta N}{2 \pi \mu k k'} \right) \]

**Relation between radius and hidden degree and radius and number of nodes (sparse network)**

\[ r = R - \frac{2}{\zeta} \ln \left[ \frac{\kappa}{\kappa_0} \right] \]
\[ N = ne^{R/\zeta} \]

**H2 Einstenian** (purely geometric)

\[ \hat{p} \left( e^{\frac{\zeta}{2} (r+r') + \frac{2}{\xi} \ln \frac{\Delta \theta}{2} - R) \right) \]
\[ \hat{p} \left( e^{\frac{\zeta}{2} (x-R)} \right) \]
\[ \hat{p}(z) \begin{cases} 
  cte & z \ll 1 \\
  0 & z \gg 1 
\end{cases} \]

\[ p(x) = \Theta(R-x) \]

nodes at hyperbolic distances smaller than \( R \) become connected

---

The inverse problem

Mapping a network to the S1 space

For each node, one finds the coordinates expected degree and the angular position on the circle that maximize the congruency between the network and the map the likelihood that the observed topology has been produced by the model

\[ L = \prod_{i<j} p(x_{ij})^{a_{ij}} [1 - p(x_{ij})]^{1-a_{ij}} \]

- Seed: subgraph of hubs disconnected and positioned at random
- Layers of lower degree nodes are sequentially added:
  - For each node, the initial position is calculated by maximizing its local contribution to the likelihood assuming old nodes are at fixed positions (considering only connections to old nodes)
  - Once all nodes in a layer are given initial positions, the positions of all old nodes plus those in the new layer are revised to improve the maximization of their local contribution to the likelihood (considering all connections)
- For low degree nodes in big networks, only first substep above (to avoid \(N^3\))

S1xS1 model for bipartite networks
The inverse problem

Validation on synthetic networks

Inferred angular coordinates vs the real ones for a networks generated with the S1xS1 model (same parameters as the real E. coli metabolism)

Distances are binned, and for each bin the ratio of the number of connected node pairs to the total number of node pairs falling within this bin is shown (calibration).
The Internet

Self-organized global system of communication: computer networks that interconnect using specific protocols (the TCP/IP internet protocol suite) specifying how data should be formatted, addressed, transmitted, routed and received.

Internet at the AS level

AS: collection of connected Internet Protocol (IP) routing prefixes under the control of one or more network operators that presents a common, clearly defined routing policy to the Internet.

Internet ≠ WWW
The Internet

AS topology, CAIDA data

23752 Ass, 58416 links

$k = 4.92$
$k_{\text{max}} = 2778$
$C = 0.61$
$\gamma = 2.1$

- Size of AS proportional to the logarithm of their degrees.
- Only ASs with a degree above 3
- Only the connections with $p(x) > 0.5$
- Font size of the country names proportional to the logarithm of the number of ASs in the country.
- Only the names of countries with more than 10 ASs are included.

M. Boguñá, F. Papadopoulos, D. Krioukov
Nature Communications 1, 62 (2010)
Distances are binned, and for each bin the ratio of the number of connected AS pairs to the total number of AS pairs falling within this bin is shown.

With greedy routing in the Internet map global topology knowledge is not needed and Internet communication protocols can be redesigned to avoid overheads and scalability problems.
Metabolic networks

Small-worlds


Heterogeneous degree distributions

High levels of clustering

Modular and hierarchical organization


Bipartite E. coli

Bowtie global connectivity structure

iAF1260 version of the K12 MG1655 strain of E. coli and human cell metabolism, both provided in the BiGG database

E. coli: 1512 reactions and 1010 metabolites, $\gamma=2.65$

human metabolism: 2201 reactions and 1482 metabolites, $\gamma=2.55$
Metabolic networks

Cartographic maps of metabolism


Metabolism and bipartite S1xS1 are extremely congruent
Metabolic networks

Localized vs transversal pathways

Some specific pathways appear concentrated over narrow sectors of polar angles, while more transversal ones are widespread over the circle. This points to a diversity of pathway topologies, with some of them displaying groups of densely interconnected reactions while some others evidencing a much more weakly connected internal structure.

Modularity of pathways
Metabolic networks

Pathways crosstalk

**Crosstalk:** absolute measure of the strength of the interaction between a pair of pathways – probability of connection with overlapping metabolites

\[ XT_{PaPb} = \sum_{j \in P_a} \sum_{j' \in P_b} \sum_{i \in \nu} (p(x_{ij}) + p(x_{ij'})) \] observed links

Very dense matrices that we filter using the disparity filter

Growing networks

<table>
<thead>
<tr>
<th></th>
<th>Topological models (no clustering)</th>
<th>Geometric models (clustering emerges)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equilibrium</strong></td>
<td><strong>Configuration Model</strong></td>
<td><strong>S₁/H₂</strong></td>
</tr>
<tr>
<td></td>
<td>Maximally random given (P(k))</td>
<td>Maximally random given (P(k)) and underlying geometry</td>
</tr>
<tr>
<td><strong>Growing</strong></td>
<td><strong>Preferential Attachment (PA)</strong></td>
<td><strong>GS₁/GH₂</strong></td>
</tr>
<tr>
<td></td>
<td>Barabási-Albert</td>
<td>PA emerges</td>
</tr>
</tbody>
</table>

Popularity is attractive but so is similarity + time
Growing networks

Popularity vs similarity in growing networks,
F Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguñá, and D. Krioukov

Popularity x Similarity optimization

(1) initially the network is empty

(2) a new node appears at a birth time $t$ and at a random angular position on the circle

(3) the new node $t$ connects to a subset of $m$ existing nodes $s$, $s < t$, with the smallest values of the product $s \theta_{st}$, where $\theta_{st}$ is the angular distance between $s$ and $t$

If radius $r_t = \ln(\text{birth time})$, PxF S minimization is the same as minimizing the hyperbolic distance between nodes (connecting to those within distance $R_t \approx r_t$)
Growing networks

The probability that an existing node of degree k attracts a connection is the same as in PA (proportional to degree) degree distributions in GH2 and PA are the same power laws

However, the probability of connection between nodes as a function of the hyperbolic distance is different in PA clustering asymptotically zero whereas it is strong in GH2

Popularity vs similarity in growing networks,
F Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguñá, and D. Krioukov
Growing networks

The strength of clustering and the exponent of the power-law can be adjusted to arbitrary values via model modifications:

• popularity fading (nodes drift away from the center) minimizing $s^\beta \theta_{st}$ \quad ($\gamma = 1 + 1/\beta$)

• clustering is weaken by allowing connection to more distant nodes with $p(x_{st}) = 1/[1 + \exp[(x_{st} - R_t)/T]]$ (T=0 is original model, maximum clustering)

Real networks can be mapped to their Popularity X Similarity spaces

Real networks seem to evolve as our framework predicts $\rightarrow$ link prediction....

Popularity vs similarity in growing networks,
F Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguñá, and D. Krioukov
## Refs

| S1 – Self-similarity | M. A. Serrano, D. Krioukov, M. Boguñá  
PRL 100, 078701 (2008), PRL 106, 048701 (2011)  
Boguna, M; Krioukov, D; Claffy, KC  
NATURE PHYSICS 5, 74-80 (2009) |
|----------------------|----------------------------------------------------------------------------------|
| H2 – S1              | Krioukov, D; Papadopoulos, F; Kitsak, M; Vahdat, A; Boguna, M  
PHYSICAL REVIEW E 82, 36106 (2010) |
| H2 – Internet        | Boguñá, M; Papadopoulos, F; Krioukov, D  
Nature Communications 1, 62 (2010) |
| S1 – Metabolic       | M. A. Serrano, M. Boguñá, F. Sagués,  
Mol. BioSyst. 8, 843-850 (2012) |
| Growing – H2         | F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguñá, D. Krioukov  