universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Unraveling hidden ring structures in real-world networks"

verfasst von / submitted by

Markus K. Youssef, BSc

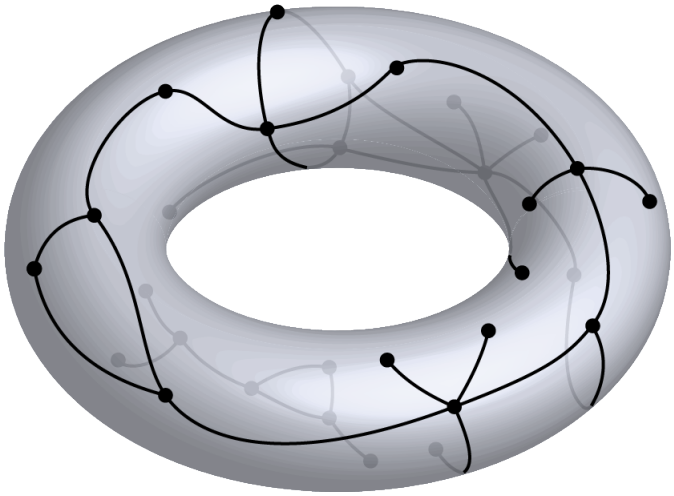
angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2019 / Vienna, 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik, geometry and topology
Betreuer / Supervisors:	Univ.-Prof. Dr. Herbert Edelsbrunner Institute of Science and Technology – Austria, Klosterneuburg

Unraveling hidden ring structures



in real-world networks

Abstract

Computational topology and network theory constitute two significant fields in nowadays science that establishes a unique bridge between mathematics and other data-driven disciplines, like bioinformatics. While examining a peculiar ring-like network pattern, it turned out that a formal explanation of that matter could only be achieved by a combination of these fields. This master thesis discloses the mathematical background more rigorously, that is needed to understand this formalism.

Zusammenfassung

In der heutigen Wissenschaft haben sich zwei prominente Gebiete, nämlich jenes der algorithmischen Topologie und jenes der Netzwerktheorie, herauskristallisiert. Diese bestreben die Kluft zwischen der reinen Mathematik auf der einen Seite und anderen, datengetriebenen Disziplinen, wie zum Beispiel derjenigen der Bioinformatik, zu überbrücken. Bei der Untersuchung eines ringartigen Netzwerks stellte sich heraus, dass eine Formalisierung dieses Sachverhaltes nur durch eine Kombination dieser beiden Gebiete möglich war. Diese Masterarbeit erläutert den dafür notwendigen mathematischen Hintergrund.

Acknowledgements

First and foremost, I want to thank *Jörg Menche* for not just having enough faith in me to pull off this ambitious project but providing me with supervision that goes far beyond any traditional kind of guidance. It is incredibly unfortunate that the system of this faculty did not allow for any sort of official recognition on that matter.
Next, I want to thank *Herbert Edelsbrunner*, who by engaging in long and intriguing discussions and scrutinizing the corresponding paper so thoroughly strongly shaped the path of this project, definitely in a better direction.
I would not have managed the last couple of month without the help of *Joel Hancock*, who was a tremendously pleasant office mate and greatly simplified our network model by spending weeks of rewriting it while being kind enough to stick to my terrible notation.
Another role as a close and helping friend had *Mateusz Piorkowski* who was always engaged in my ideas and stimulated them by constructive comments.
A crucial, but more subtle, role for this project and, really, in my life played *Sophie Wegscheider*, who was always there and supported me as much as possible.
Last, but not least, I want to thank *my family*, who provided me with the opportunity even to get this far.

Thank you all for making this happen!

Preface

The content of this thesis arose from a project at the CeMM - Research Center for Molecular Medicine of the Austrian Academy of Sciences that was lead by Jörg Menche. The main intent was to formalize the notion of a ring-structured network and to develop an understanding of its behavior.
Its origin can be traced back years ago when Berend Snijder analyzed co-regulation patterns of lipid molecules that displayed a pronounced circular organization in its network representation. The paper for the corresponding biological mechanism was published in 2015. Intrigued by the structure of this particular network, a pet-project emerged, intending to quantify the 'ringy-ness' of any given network. After countless failed attempts to characterize this matter within the scope of network theory alone, we concluded that tools from computational topology could be exploited to tackle this problem. This insight resulted in a paper that will be submitted in the near future.
In this thesis, I will cover the mathematical background that was necessary to develop the theory that gave rise to that paper.

Contents

Introduction

One central problem in computational topology is how to retrieve topological information from incomplete or noisy data. *Persistent homology* focuses, as the name suggests, on retrieving objects that, in lights of the theory of algebraic topology, are known as *homology groups*. These objects are one possible mathematical formulation of the concept of "holes" at different dimensions. The first homology groups, for example, aims to formalize the notion of a "tunnel". In this context, a ring can be seen as an object whose first homology group has rank 1.
The overall idea of persistent homology is to look at a whole filtration of topological spaces and to identify homology groups at different steps of this filtrations, leading to the concept of *persistence*. Homology groups that persist longer are believed to be more likely a feature of some underlying geometry.

While topology fundamentally deals with uncountable data, finite graphs, also called *networks*, are inherently countable objects and can hence be seen as "incomplete" in the eyes of a topologist. Accordingly, the question of whether a given system is ring-like in its network representation naturally leads to an application of persistent homology.
Once the concept of ring-networks is established, the next aim was to develop a new *network model* to gain a more in-depth understanding of the mechanisms that lead up to the emergence of ring-structures. To that end, we introduce a generic system setup and demonstrate how classical network models fit into this framework. We then proceed to the proposed network model. A detailed analysis of each component in the network model can be found in the paper this thesis is based on.

1 Computational topology

One Ring to rule them all,
One Ring to find them,
One Ring to bring them all
and in the darkness bind them.

J. R. R. Tolkien,
The Lord of the Rings

For the last century, the field of algebraic topology developed many rigorous concepts of how to analyze the topological structure of a given space. One idea that branched of as a whole field on its own right is known as (simplicial) homology, a theory that aims to formalize the notion of a "hole" is [5]. Towards the end of the century, new ideas started to emerge to deal with the problem of having incomplete data of geometric objects, for example, when only sampled measurements are available rather than a full description of the underlying geometry. One cornerstone of this advancement was the approach of what is now known as *persistence* [23, 11, 7]. In this chapter, we will only explain the very basics of this idea, just enough to develop an intuition the construction of the ring score at the end of this thesis. For further readings on computational topology, we refer to the book by Edelsbrunner and Harer [8].

1.1 Simplicial complexes

The underlying assumption in the theory of simplicial homology is that one can represent, or at least approximate, an object of interest as a combinatorial data structure known as a simplicial complex.

Definition 1.1 (Simplicial complex)**.** Let V be a finite set. A set of subsets $\mathcal{K} \subset \mathfrak{P}(V)$ is called an *(abstract) simplicial complex* on the *vertex set* V if the following two conditions hold:

- $\mathcal{K}$ is a cover, i.e. $\bigcup_{\sigma \in \mathcal{K}} \sigma = V$
- $\mathcal{K}$ is downwards closed, i.e. $\forall \tau \in \mathcal{K} : \sigma \subset \tau \Rightarrow \sigma \in \mathcal{K}$

Elements of a simplicial complex are called *simplices*. Simplicial complexes are naturally equipped with a partition into simplices of the same size. More precisely, an element of a simplicial complex on $n+1$ vertices is called an *n-dimensional simplex* or *n-simplex* for short. Note, that we will, by abuse of notation, make no distinction between a 0-simplex, $\{v\}$, and its underlying vertex v.

This partition gives rise to a sequence of vector spaces, collectively known as an (induced) chain complex.

Definition 1.2 (Simplicial Chain)**.** An arbitrary set of n-simplices is called an *n-chain with values in $\mathbb{Z}_2$*. We will denote the set of all n-chains with values in $\mathbb{Z}_2$ by C_n. Given two n-chains $c_1, c_2 \in C_n$ we define their *symmetric difference* $c_1 + c_2$ to be

$$c_1 + c_2 = \{\sigma \in c_1 \cup c_2 \mid \sigma \notin c_1 \cap c_2\} .$$

This defines an idempotent group, $(C_n, +, \emptyset)$, on the set of all n-chains. Hence C_n embodies a natural vector space structure over $\mathbb{Z}_2$. In this context, one-element chains, i.e., n-simplices, constitute a natural choice of a basis for C_n.

As suggested by the name chain "complex", the chains, C_n, define not merely a sequence of disjoint vector spaces but are accompanied by linear transformations, known as boundary maps.

Definition 1.3 (Boundary maps)**.** Given an n-simplex σ, any k-simplex contained in σ is called a *k-face of σ*. We define the *n-th boundary map*, denoted by ∂_n, as a function on n-chains that sends each n-simplex to all of its n-faces considered as an $(n-1)$-chain. Since n-simplices constitute a basis for C_n this map uniquely extends to a well-defined linear map

$$\partial_n : C_n \to C_{n-1} : \{\sigma_1, \ldots, \sigma_k\} \mapsto \sum_{i=1}^{k} \partial_n(\sigma_i) .$$

Chains that lie in the image of boundary maps are called *boundaries*,

$$B_n = \operatorname{im} \partial_{n+1} \subset C_n ,$$

while elements that are sent to zero are called *cycles*,

$$Z_n = \ker \partial_n \subset C_n .$$

The fact that every boundary is a cycle is *the* defining property of homology theory.

Theorem 1.4 (Fundamental lemma of homology)**.** The sequence $(C_n, \partial_n)_{n\in\mathbb{N}}$ indeed constitutes a chain complex in the algebraic sense. In other words, we have $\partial_n \partial_{n+1} = 0$ and hence $B_n \subset Z_n \subset C_n$ for all dimensions $n \in \mathbb{N}$.

Proof. We will show that $\partial_n \partial_{n+1} \sigma = 0$ for any $(n+1)$-simplex σ. By definition, the boundary $\partial_{n+1}\sigma$ consists of all n-faces of σ. Every $(n-1)$-face of σ belongs to exactly two n-faces, so $\partial_n(\partial_{n+1}\sigma) = 0$. Since $\partial_n \partial_{n+1}$ is a linear map that vanish on a basis, it must be the zero map. □

A chain complex $(C_\bullet, \partial_\bullet)$ is called *exact* if $B_n = Z_n$, i.e. if all cycles are also boundaries. The property of being exact formalizes the notion of having "no holes". If we think of cycles as being "openings" in the simplicial complex, exactness implies that every such opening arises as the boundary of some higher-dimensional object, similar to the opening at the cuff of a sock being just the boundary of the sock itself. In the unfortunate case of a sock exhibiting an additional opening at its toe, each individual opening cannot emerge as the boundary of the sock without including the other one. Thus, one considers the sock to have at least one hole. With that analogy in mind, homology groups describe the appearance of holes by measuring "how far" a chain complex is from being exact.

Definition 1.5 (Homology group)**.** The *n-th homology group* of a simplicial complex with boundaries B_n and cycles Z_n is defined to be the quotient group

$$H_n = Z_n / B_n \,.$$

Since Z_n and B_n are vector spaces over $\mathbb{Z}_2$ the homology group H_n additionally inherits a vector space structure. The dimension of this vector space is called the *n-th Betti number*, $\beta_n = \dim H_n$.

Betti numbers ought to be thought of as the amount of "n-dimensional holes" of a simplicial complex.

1.2 Persistent homology

The idea of persistence is that no single simplicial complex is good enough to approximate a geometric structure, given a set of sample points. Instead, one is interested in a whole family of approximations. We will restrict ourselves to a special construction that parametrizes this family by a single parameter $r \in \mathbb{R}_{\geq 0}$.

Definition 1.6 (Vietoris-Rips complex)**.** Given a finite metric space (M, d) and a radius $r > 0$, the simplicial complex $\mathcal{V}_r$ defined by

$$\mathcal{V}_r := \{\sigma \subset M \mid d(x,y) \leq 2r \; \forall x, y \in \sigma\} \,,$$

is called the *Vietoris Rips complex of M at radius r*.

This construction has the convenient property that it is independent of the embedding the sample points, i.e., only the pairwise distances are relevant, and that it constitutes a filtration of simplicial complexes in the sense that $\mathcal{V}_r \subset \mathcal{V}_{r'}$ whenever $r \leq r'$. This allows for a natural identification of homology groups at different steps of the filtration.

Definition 1.7 (Persistence)**.** For $r_i \leq r_j$ there are obvious inclusions of the corresponding complexes $\iota_{ij} : \mathcal{V}_{r_i} \hookrightarrow \mathcal{V}_{r_j}$, which further induce homomorphisms on their corresponding homology groups $f_n^{i,j} : H_n(\mathcal{V}_{r_i}) \to H_n(\mathcal{V}_{r_j})$ for all $n \in \mathbb{N}$. Since M is finite, there is only a finite amount of radii, call them $r_0 \leq \cdots \leq r_N$, for which $H_n(\mathcal{V}_{r_i-\epsilon}) \to H_n(\mathcal{V}_{r_i})$ is **not** an isomorphism for all $\epsilon > 0$ sufficiently small. We say a homology class $\gamma \in H_n(\mathcal{V}_{r_i})$ is *born at* r_i, if $\gamma \notin \operatorname{im} f_n^{i-1,i}$. Furthermore, if γ is born at r_i then it *dies entering* r_j, if it merges with an older class as we go from $\mathcal{V}_{j-1}$ to $\mathcal{V}_j$, that is, if $f_n^{i,j-1}(\gamma) \notin \operatorname{im} f_n^{i-1,j-1}$ but $f_n^{i,j}(\gamma) \in \operatorname{im} f_n^{i-1,j}$. If a homology class is born at r_i and dies entering r_j, then we define its *persistence* to be $\operatorname{pers}(\gamma) = r_j - r_i$.

There are several approaches to keep track of all relevant persistence data in an organized fashion. One wildly used representation is the so-called persistence diagram.

Definition 1.8 (Persistence diagram)**.** Let $\mathcal{V} = (\mathcal{V}_{r_i})_{r_i \in R}$ be a filtration of simplicial complexes indexed by a countable set of real numbers $R \subset \mathbb{R}$. We define its n-th *persistence diagram*, $\operatorname{dgm}_n(\mathcal{V})$, to be the collection of all birth-death pairs, (r_i, r_j), for all n-th homology classes appearing in the filtration. It is often convenient to add the "diagonal with infinite multiplicity" to the diagram, meaning that the pair (r, r) appears infinitely often in the diagram for all $r \in R$. Caution has to be taken in that homology classes might not die at all. Hence, points in the diagram take values in the extended real line $\overline{\mathbb{R}}_{\geq 0} = \mathbb{R}_{\geq 0} \cup \{\infty\}$. A filtration $\mathcal{V}$ arising from a Vietoris-Rips complex construction is completely determined by the pairwise distance matrix, D, of the underlying finite metric space. In this case we will use the unambiguous notation $\operatorname{dgm}_n(D) = \operatorname{dgm}_n(\mathcal{V})$.

For our purposes, we are only concerned about the (size of the) persistences itself and do not need the full information of the persistence diagram. We thus reduced the persistence date further to a normalized decreasing sequence of real numbers.

Definition 1.9 (Persistence sequence)**.** Given a persistence diagram $\operatorname{dgm}(\mathcal{V})$ for each birth-death pair $(r_{\text{birth}}, r_{\text{death}}) \in \operatorname{dgm}(\mathcal{V})$ its corresponding persistence is given by $\operatorname{pers} = r_{\text{death}} - r_{\text{birth}}$ its corresponding persistences are given by. We will order them by size and divide them by the maximal persistence

such that $1 = p_0 \geq p_1 \geq p_2 \dots$ with $p_i = \mathrm{pers}_i \,/\, \mathrm{pers}_{\max}$. We will refer to the descending sequence $\mathbf{p} = (p_0, p_1, p_2, \dots)$ padded with infinitely many zeros at the end as the *(normalized) persistence sequence* of $\mathrm{dgm}(\mathcal{V})$.

2 Elementary probability theory

> If you like it, then you shoulda put a ring on it.
>
> Beyoncé,
> Single Ladies

First, we turn our attention to basic concepts of measure and probability theory. Most of the probability spaces in this work are countable or, in fact, even finite, making the heavy machinery of measure theory obsolete. However, abstract probability theory provides a powerful and unifying language, allowing a clean and general formulation of our network model proposed in chapter 4. For further readings on probability theory, we refer to the book by Klenke [14].

Definition 2.1 (Sample space)**.** Let Ω be a set and $\mathcal{F} \subset \mathfrak{P}(\Omega)$ be a subset of the power set of Ω. The pair $(\Omega, \mathcal{F})$ is called a *sample space* or a *measurable space* if $\mathcal{F}$ is a σ-algebra on Ω, i.e. if the following three conditions hold:

1. $\mathcal{F} \subset \mathfrak{P}(\Omega)$ and $\mathcal{F} \neq \emptyset$
2. If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$
3. If $A_k \in \mathcal{F}$ for all $k \in \mathbb{N}$, then $\bigcup_{k\in\mathbb{N}} A_k \in \mathcal{F}$

By abuse of notation, we will refer to the underlying set Ω itself as the sample space and call elements of $\mathcal{F}$ *events*.

A sample space is just a "preparation" for a probability space. It turns out that a "nice" notion of a probability measure usually comes with a restriction of the underlying σ-algebra $\mathcal{F}$. Yet, in case Ω is countable, there are almost no problems with allowing every subset of Ω to be an event, i.e., $\mathcal{F} = \mathfrak{P}(\Omega)$. Hence, we will always assume $\mathcal{F} = \mathfrak{P}(\Omega)$ whenever Ω is countable.

Definition 2.2 (Probability space)**.** Let $(\Omega, \mathcal{F})$ be a sample space. A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ is called a *probability measure* on $(\Omega, \mathcal{F})$ if the following two conditions hold:

1. $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.
2. If $A_k \in \mathcal{F}$ for all $k \in \mathbb{N}$ are pairwise disjoint, then

$$\mathbb{P}\left(\dot{\bigcup_{k\in\mathbb{N}}} A_k\right) = \sum_{k\in\mathbb{N}} \mathbb{P}(A_k)\,.$$

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is then called a *probability space* or more ambiguously a *probability distribution*.

Typically, the probability space itself is a comparatively complicated object. To circumvent that we reduce the abstract probability space to a probability space on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the *Borel algebra*, i.e. the σ-algebra generated by the set of all open intervals of $\mathbb{R}$.

Definition 2.3 (Random variable)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A function $X : \Omega \to \mathbb{R}$ such that $X^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}$ is called a *random variable*. Every random variable X induces a probability measure $\mathbb{P}_X$ on $(\mathbb{R}, \mathcal{B}, \mathbb{P})$, called the *induced probability measure*, via $\mathbb{P}_X[B] = \mathbb{P}[X \in B]$ for all $B \in \mathcal{B}$, where we used the notation $[X \in B]$ instead of $X^{-1}(B)$. Given a probability measure $\mathbb{P}_X$ on $(\mathbb{R}, \mathcal{B})$ it can be fully described via its *cumulative distribution function*

$$F_X(t) = \mathbb{P}_X\big[(-\infty, t]\big] = \mathbb{P}[X \leq t]\,,$$

where, again, $[X \leq t]$ stands for $X^{-1}((-\infty, t])$. In case X only takes countable many values we can further define the *probability mass function* to be

$$p_X(t) = \mathbb{P}_X\big[\{t\}\big] = \mathbb{P}[X = t]\,.$$

If, on the other extreme, F_X is absolutely continuous we define the *probability density function* to be

$$f_X(t) = F_X'(t)\,,$$

for all $t \in \mathbb{R}$ where F_X is differentiable, and leave F_X undefined otherwise.

A useful tool to calculate probabilities is to condition on events that already happened or at least we know how likely it is that they will happen. Understanding conditional probabilities is crucial for an understanding of a Markov chains and random walks.

Definition 2.4 (Conditional probability and independence)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $A, B \in \mathcal{F}$ two events. The *conditional probability* of A given B is defined as

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}\,.$$

We say that A is independent of B if conditioning on it does not change its probability. This is equivalent to saying that the probability of the mutual event is the product of the probability of the individual events, i.e.

$$\mathbb{P}[A \mid B] = \mathbb{P}[A] \Leftrightarrow \mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]\,.$$

Given a family of events, $(A_i)_{i\in I}$ we say they are *mutual independent* if

$$\mathbb{P}\left[\bigcap_{i\in I} A_i\right] = \prod_{i\in I} \mathbb{P}[A_i]$$

holds. Note however that pairwise independence does **not** imply mutual independence. Lastly, we mention that given a discrete random variable X the term $[X = x] = X^{-1}(x) \in \mathcal{F}$ is simply an event and so expressions like $\mathbb{P}[A \mid X = x]$ are well-defined.

Let us assume, we want to find out if a given die is biased or not. In the light of statistical inference, we aim to identify, if the discrete uniform distribution on the set of pips explains our observed data the best **out of a predefined family of probability distributions**. As the formating suggests, the second part of the sentence is crucial for the definition of a statistical model itself.

Definition 2.5 (Statistical model)**.** Let $(\Omega, \mathcal{F})$ be a sample space and Θ a set. A set of probability measures $\mathcal{P}_\Theta = \{\, \mathbb{P}_\theta : \theta \in \Theta\}$ over $(\Omega, \mathcal{F})$ is called a *statistical model* on $(\Omega, \mathcal{F})$ with *parameter space* Θ. If the assignment $\theta \mapsto \mathbb{P}_\theta$ is injective we say the parametrization is *identifiable*. If $\Theta \subset \mathbb{R}^d$, i.e. if Θ is finite dimensional, then the statistical model is called a *parametric model of dimension d*.

In this work, all statistical models are assumed to be identifiable and parametric. Furthermore, instead of viewing $\theta \in \Theta$ as a single parameter of dimension d it is better to think of $\theta = (\theta_1, \ldots, \theta_d)$ as being d separate parameters. For further readings on statistical modeling, we refer to the survey by McCullagh [17].

3 Network theory

Der echte Ring,
Vermutlich ging verloren.

Gotthold Ephraim Lessing
Nathan der Weise

A graph is the mathematical abstraction of how to deal with "relational data". It contains *vertices* and *edges* that do not carry any meaning by themselves. Although, there is no clear distinction between a graph and a network we will encourage the reader to think of networks as being "graphs that represent something". In this sense, graph theory considers graphs to be interesting objects on their own right, while network theory focuses on model fitting and discovering beneficial summary statistics instead. For further readings on graph theory we refer to the book by Diestel [6] and for network theory to the book by Newman [20].

3.1 Network properties

A key philosophy of how to approach large data sets is to calculate descriptive quantities known as *statistics*. In case the data consists of real numbers, well known statistics are the *mean*, *mode*, *variance* and so on. Now, a single network can be considered as a large data set. We aim to find important random variables to convert the complicated network into a collection of real numbers and reduce the problem of "understanding the graph" into something we seem to be more familiar with. In the light of network science, these statistics are more commonly known as *network properties*.

3.1.1 Elementary network statistics

There are many ways to formalize the notion of a graph, and not all of them are necessarily equivalent. For example, the objects that we call *graphs* here are more unambiguously known as *simple graphs*. Note, however, that the way we define graphs is somewhat unusual but fits better in our framework.

Definition 3.1 (Graph)**.** Let G be a simplicial complex. If every simplex has dimension at most 1, then G is called a *graph* or a *network*. In this context, a 0-simplex is called a *vertex*, or a *node*, and a 1-simplex is called an *edge*, a *link* or an *interaction*. We can thus identify a graph with the pair (V, E), where V is the underlying set of all vertices and E the set of all edges.

If two vertices constitute an edge, we will call these vertices *adjacent*. We emphasize again that, as in the definition of a simplicial complex, we will tacitly identify a 0-simplex $\{v\}$ with its underlying vertex v.

Probably the most immediate description of a node is its degree. Consequently, the collection of all degrees in a given graph might give us some insights into its structure.

Definition 3.2 (Degree distribution)**.** Let $G = (V, E)$ be a graph and $v \in V$ a vertex. The *degree of* v, denoted by $\deg v$, is the number of edges that are incident to v, i.e.

$$\deg v = \#\{e \in E \mid v \in e\} .$$

The function $\deg : V \to \mathbb{N}$ is called the *degree sequence* of G. It can be easily shown, e.g. by using double counting, that for $M = |E|$ we have

$$\sum_{v \in V} \deg v = 2M .$$

Let further $P(k)$ be the number of nodes in G with degree k. The function $P : \mathbb{N} \to \mathbb{N}$ is called the *degree distribution* of G. It can easily be shown that for $N = |V|$ we have

$$\sum_{k \in \mathbb{N}} P(k) = N .$$

Let us motivate how these concepts fit into the framework of calculating statistics. The key to translate graph theoretic properties into probabilistic ones is to assume a uniform probability measure $\mathbb{P}$ on the vertex set V, i.e. $\mathbb{P}[\{v\}] = \frac{1}{N}$ for all $v \in V$. Then deg can be seen as a random variable on the probability space $(V, \mathfrak{P}(V), \mathbb{P})$. As a random variable it induces a probability measure, $\mathbb{P}_{\deg}$, on $\mathbb{N}$ with probability mass function

$$p_{\deg}(k) = \mathbb{P}[\deg = k] = \frac{P(k)}{N} .$$

Hence, the degree distribution of a graph indeed becomes a probability distribution and we can make statements like "The degree distribution of a graph is binomial." Taking this even further, we can represent the *average degree* of a graph as the expected value of the random variable deg, i.e.

$$\mathbb{E}[\deg] = \sum_{k \in \mathbb{N}} k\mathbb{P}[\deg = k] = \frac{1}{N} \sum_{k \in \mathbb{N}} kP[k] \overset{!}{=} \frac{1}{N} \sum_{v \in V} \deg v = \frac{2M}{N} ,$$

where the equation with the exclamation mark can be seen by grouping together vertices of the same degree in the right sum. Note, that we can obtain the *density* of a graph from the average degree via

$$\varrho = \frac{M}{\binom{N}{2}} = \frac{\mathbb{E}[\deg]}{N-1}\,.$$

Another crucial concept in network theory is to quantify how far two nodes are from each other. Indeed, every graph carries a natural metric structure known as the shortest-path-length metric.

Definition 3.3 (Shortest path length)**.** Let G be a graph. A sequence of pairwise distinct vertices $\gamma = (\gamma_0, \ldots, \gamma_n)$ for which (γ_i, γ_{i+1}) is adjacent for all $i = 0, \ldots, n-1$ is called a *path*. The number of edges in this sequence, n, is called the *path length of* γ, denoted by $L_{\mathrm{SPL}}(\gamma)$. Given two vertices in a graph u and v we can look at the set of all paths with *starting point* $u = \gamma_s$ and *end point* $v = \gamma_t$. The infimum of the length over all these paths is called the *shortes path length* of u and v, denoted by $d_{\mathrm{SPL}}(u, v)$. Formally, we can write

$$d_{\mathrm{SPL}}(u, v) = \inf\left\{ L_{\mathrm{SPL}}(\gamma) \mid \gamma_s = u, \gamma_t = v \right\}.$$

A graph for which $d_{\mathrm{SPL}}(u, v) < \infty$ for all $u, v \in V$ is called *connected*.

For the sake of completeness we will also cover the notion of a clustering coefficient, probably first introduced by Luce and Perry [16] and popularized by Watts and Strogatz when they presented their small-world model [24]. Intuitively, it measures how close the neighbourhood of a node is from being a complete subgraph.

Definition 3.4 (Clustering coefficient)**.** Let $G = (V, N)$ be a graph. The *local clustering coefficient* $C(v)$ of a vertex v is given by the fraction of edges between adjacent vertices of v and the number of edges that could possibly exist between them. Formally, we write

$$C(v) = \frac{\#\left\{ \{u, w\} \in E \;\middle|\; \{u, v\}, \{v, w\} \in E \right\}}{\binom{\deg v}{2}}\,.$$

3.1.2 Random walk distance

Even though the shortest path length embodies a natural choice for a metric to any given network, it turns out that this metric is to "coarse" for many applications due to the small world property of real life networks. The topology of a graph can be elegantly analyzed with the use of random walks. A

random walk on a graph can be thought of as a being, jumping from vertex to any adjacent vertex using the corresponding edge. At any given step, the probability of taking a particular edge is proportional to the degree of the current vertex, i.e., the number of possible edges to choose from. We will exploit this tool to define a metric on any given network that is more fine-grained than the shortest path length.

Definition 3.5 (Random walk)**.** Let $G = (V, E)$ be a graph and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. A sequence of V-valued random variables $X_i : \Omega \to V$ is called a *Markov chain* if

$$\mathbb{P}[X_{n+1} = v_{n+1} \mid X_n = v_n, \ldots, X_0 = v_0] = \mathbb{P}[X_{n+1} = v_{n+1} \mid X_n = v_n]\,,$$

for all $n \in \mathbb{N}$ and $v_i \in V$, for which $\mathbb{P}[X_0 = v_0, \ldots, X_n = v_n] > 0$. In this case the matrix $\mathbf{W} = (w_{ij})_{i,j \in V}$, given by the entries $w_{ij} = \mathbb{P}[X_{n+1} = j \mid X_n = i]$ for any choice of $n \in \mathbb{N}$, is well defined and called the *transition matrix* of the Markov chain. A Markov chain is called a *random walk* on G *starting at* s, if

$$\mathbb{P}[X_0 = s] = 1 \text{ and } \mathbb{P}[X_{n+1} = v_{n+1} \mid X_n = v_n] = \frac{1}{\deg v_n}\,,$$

whenever v_{n+1} and v_n are adjacent, and zero otherwise. A random walk is said to be an *absorbing random walk with absorbing state* t, if the above equation holds for all $v_i \neq t$. For $v = t$ we replace the above condition with $\mathbb{P}[X_{n+1} = t \mid X_n = t] = 1$. Given a random walk with absorbing state t we define the *fundamental matrix* to be $\mathbf{N}(t) = (\mathbf{I} - \mathbf{Q}(t))^{-1}$, where $\mathbf{Q}(t)$ is the transition matrix with the t-th row and column removed and $\mathbf{I}$ is the identity matrix.

The fundamental matrix is mainly a tool to calculate characteristics of the random walk. One of these characteristics we want to draw attention to is the expected number of visits for a given node.

Theorem 3.6 (Expected number of visits)**.** Let G be a graph and fix a node t. Then the expected number of visits to any node $j \neq t$ for an absorbing random walk starting at $s \neq t$ with absorbing state t is given by the (s, j)-th entry of the fundamental matrix $\mathbf{N}(t)$.

Proof. The probability of transitioning from any node i to j in exactly k steps is given by $w_{ij}^{(k)}$, i.e. the (i, j)-entry of the k-th power of the transition matrix W. Because of

$$w_{ij}^{(2)} = \sum_{l \in V} w_{il} w_{lj} = \sum_{l \in V \setminus \{t\}} w_{il} w_{lj} + w_{it} w_{ti} \overset{(w_{ti}=0)}{=} \sum_{l \in V \setminus \{t\}} w_{il} w_{lj} = q_{ij}^{(2)}(t)\,,$$

we have $w_{ij}^{(k)} = q_{ij}^{(k)}(t)$ for all $i, j \neq t$ and for all $k \in \mathbb{N}$, taking into account an appropriate labeling of the matrix $\mathbf{Q}(t)$. Hence, the expected number of visits to a transient state j starting from a transient state i is given by the (i, j)-entry of

$$\sum_{k\in\mathbb{N}} \mathbf{Q}^k(t) = (\mathbf{I} - \mathbf{Q}(t))^{-1} = \mathbf{N}(t)\,.$$

□

Next, we focus on a concept called *centrality measure* or *centrality* for short. A centrality measure gauges how "important" a node or an edge is for the structure of the whole network. The following centrality measure is a random walk based edge centrality introduced by Newman [19]. Because of the similar mathematical expressions of random walk characteristics and electrical network characteristics, many terms have their origin from electrical circuits.

Definition 3.7 (Current flow centrality)**.** Let $G = (V.E)$ be a network. For a fixed source-target pair (s, t) the expected number of visits to any node $i \neq t$ is given by $N_{si}(t)$. Hence, the expected number of times the walker (then) walks from node i to any fixed adjacent node j is given by $U_i^{st} = N_{si}(t)/\deg i$. Because we know that the walker will never visit any of its neighbours once it reaches t we have $U_t^{st} = 0$. We will refer to the collection $\mathbf{U}^{st} = (U_v^{st})_{v\in V}$ as the *voltage vector*. The (s, t)-*net flow* along the edge $\{i, j\}$ is defined to be $|U_i^{st}–U_j^{st}|$, i.e. the absolute difference of how often the random walker crosses this edge in one direction versus the other direction. The *net flow centrality* is defined to be the average (s, t)-net flow

$$\tau_{ij} = \frac{2}{|V|(|V|-1)} \cdot \sum_{s<t} |U_i^{st} - U_j^{st}|\,. \qquad (*)$$

To extend the current flow centrality to a metric on the whole network, we first need to check if at least locally it satisfies the right conditions.

Theorem 3.8 (Triangular inequality)**.** Let $\big(\{i, j\}, \{i, k\}, \{j, k\}\big)$ be a triangle of edges in a network. Then the triangular inequality for their corresponding current flow centralities holds, i.e.

$$\tau_{ik} \leq \tau_{ij} + \tau_{jk}\,.$$

Proof. For each pair of nodes (s, t) the inequality

$$|U_i^{st} - U_k^{st}| = |U_i^{st} - U_j^{st} + U_j^{st} - U_k^{st}| \leq |U_i^{st}–U_j^{st}| + |U_j^{st}–U_k^{st}|\,,$$

holds. Hence the inequality holds for each term in the sum of equation $(*)$. The claim now follows from the invariance of inequalities under multiplication by a positive constant. □

It might seem unintuitive at first to equate a measure of importance to a measure of distances. However, if we think of a traffic network, it makes sense to say that a long road is more influential to the structure of the network than a short side street. In that spirit, we have all the tools to finally define a random walk based distance measure on any given graph.

Definition 3.9 (Current flow distance)**.** Let G be a connected network and $\gamma = (\gamma_0, \dots, v_n)$ a path in G. We define the *current flow length* of the path γ to be the sum of all the current flow centralities of the edges appearing in γ, i.e.

$$L_{\text{flow}}(\gamma) = \sum_{i=0}^{n-1} \tau_{v_i v_{i+1}} .$$

The *current flow distance* of two arbitraty nodes u and v is now defined to be the minimal current flow length between u and v, i.e.

$$d_{\text{flow}}(u, v) = \inf \left\{ L_{\text{flow}}(\gamma) \mid \gamma_s = u, \gamma_t = v \right\} .$$

Note that in general, the path of minimal current flow does not necessarily coincide with the respective (unweighted) shortest path. Nonetheless, the triangular inequality ensures that an already existing edge (u, v) will not be 'overwritten' by $d_{\text{flow}}(u, v)$.

3.2 Random graphs

Given a network constructed from a set of relational data, we wish to gain more insights into the organizational nature of this data. One approach is to propose a generating mechanism based on a principle of interest to produce a synthetic network that ideally mimics the original network of interest. Such synthetic networks are known as network models.

Definition 3.10 (Network model)**.** Let $[N] = \{1, \dots, N\}$ be a fixed set on N elements and denote the set of all graphs with vertex set $V = [N]$ by Ω_N and define further $\Omega = \dot{\bigcup}_{N \in \mathbb{N}} \Omega_n$. A statistical model on $(\Omega, \mathfrak{P}(\Omega))$ is called a *network model*.

In this work, the number of nodes is always assumed to be a parameter of the network model. More formally, all network models presented here have a parameter space of the form $\Theta = \mathbb{N} \times \Theta'$ with $\mathbb{P}_{(N,\Theta')}[\Omega_N] = 1$ for all

$N \in \mathbb{N}$. Hence, a network model with fixed parameter N can be regarded as a statistical model on $(\Omega_N, \mathfrak{P}(\Omega_N))$ with parameter space Θ'. In particular, the set of nodes is fixed. Thus, the probability of an edge appearing in a network model for a fixed pair of nodes is well defined.

Definition 3.11 (Interaction probability matrix)**.** For a given probability measure $\mathbb{P}$ on Ω_N, the set of all graphs with a fixed set of N nodes, the probability measure is completely described by its *interaction probability matrix* $\mathbf{X} = (X_{ij})_{i,j=1}^N$, consisting of the random variables for which $X_{ij}(G) = 1$ if the nodes $\{i, j\}$ are adjacent in G and $X_{ij}(G) = 0$ otherwise. This means that we can recover the probability of obtaining a graph G via

$$\mathbb{P}[G] = \mathbb{P}\left[\bigcap_{\{i,j\}\in E} [X_{ij} = 1] \cap \bigcap_{\{i',j'\}\notin E} [X_{i'j'} = 0]\right].$$

The diagonal of $\mathbf{X}$ can be filled with additional node attributes and will be set to zero otherwise. Note that while the X_{ij}'s individually are necessarily Bernoulli distributed, we do not assume independence and thus, cannot make general statements about a collection of X_{ij}'s. In case of mutual independence we can indeed identify each X_{ij} with its success probability $p_{ij} \in [0, 1]$ and identify the matrix $\mathbf{X}$ with $P = (p_{ij})_{i,j=1}^N$. This does not lead to a loss of information because there is no dependency structure to keep track of.

The concept of an interaction probability matrix can easily be adapted to the setting of a whole network model $\mathbb{P}_\Theta$ by parametrizing the interaction matrix $\mathbf{X}_\theta$ for each θ in the parameter space Θ. We will give some examples of the most prominent network models.

Example 3.12 (Erdős–Rényi model)**.** The *Erdős–Rényi model* is probably the oldest and most well known network model [9, 13]. It has 2 parameter and the parameter space is given by

$$\Theta = \mathbb{N} \times [0, 1].$$

Given a pair $(N, \varrho) \in \Theta$ it constructs a graph on N nodes by including each edge in the graph with probability ϱ independent from every other edge. By definition this means, that the random variables in the interaction probability matrix are mutually independent and we can hence identify $\mathbf{X} = \varrho(\mathbf{E} - \mathbf{I})$, where $\mathbf{E}$ is the all-ones matrix and $\mathbf{I}$ the identity matrix. For a given graph $G = (V, E)$ with $N = |V|$ and $M = |E|$ we can calculate its probability as

$$\begin{aligned}\mathbb{P}(\{G\}) &= \left(\prod_{\{i,j\}\in E} \mathbb{P}[X_{ij} = 1]\right) \cdot \left(\prod_{\{i',j'\}\notin E} \mathbb{P}[X_{i'j'} = 0]\right) \\ &= \varrho^M \cdot (1 - \varrho)^{\binom{N}{2} - M}.\end{aligned}$$

Hence, for $\varrho = 0.5$ we obtain the uniform distribution on the set of all graphs on N nodes.

Example 3.13 (Barabási–Albert model)**.** The *Barabási–Albert model* is a network model that aims to explain the scale-free degree distribution behaviour of a network based on a preferential attachment mechanism [3]. It has 2 parameters and the parameter space is given by

$$\Theta = \{ (N, m) \in \mathbb{N} \times \mathbb{N} \mid 0 < m < N \} .$$

Given a pair $(N, m) \in \Theta$ we construct a graph on N nodes according to a *preferential attachment process*. First, we start with a star-shaped graph on $m + 1$ nodes and call this the 1-st time step. In each new time step we add a new node u, connecting it to m existing nodes with probability proportional to its degree. More precisely, given a node v with degree $\deg_t v$ at time t, the probability of the new node, u, being connected to v is given by $\deg_t v / \sum_w \deg_t w$. Not all possible graphs can be obtained by this process. For example, the Barabási–Albert model is necessarily concentrated on the set of all connected graphs on N nodes and $m(N - m)$ edges, i.e.

$$\mathbb{P}_{N,m} \left(\left\{ G = (V, E) \in \Omega_N \;\middle|\; G \text{ connected}, |E| = m(N - m) \right\} \right) = 1 .$$

By design, this network model accumulates *hubs*, nodes with relatively high degrees, and has a scale-free degree distribution. Since the number of edges is deterministic in the model parameters, the random variables of the interaction probability matrix, X_{ij}, cannot be independent for $i, j > m + 1$.

Example 3.14 (Watts–Strogatz model)**.** The *Watts–Strogatz model* is a network model that aims to explain the small-world property of a network while maintaining a high clustering coefficient based on rewiring scheme of a regular lattice [24]. It has 3 parameters and the parameter space is given by

$$\Theta = \{ (N, k, p) \in \mathbb{N} \times \mathbb{N} \times [0, 1] \mid k \text{ is even and } k < N \} .$$

Given a tripple $(N, k, p) \in \Theta$ we first construct a regular ring lattice on N nodes, by introducing an edge $\{i, j\}$ if and only if $|i - j| \mod \left(N - \frac{k}{2}\right) \leq \frac{k}{2}$. Next, for every node i we take every edge connecting i to its $\frac{k}{2}$ "right" neighbors and rewire it with probability p by replacing $\{i, j\}$ with $\{i, l\}$ where l is chosen uniformly at random from all possible nodes while avoiding self-loops and link duplication, see Fig.3.14 for an illustration. Since the number of edges is deterministic in the model parameters, the random variables of the interaction probability matrix, X_{ij}, cannot be independent.

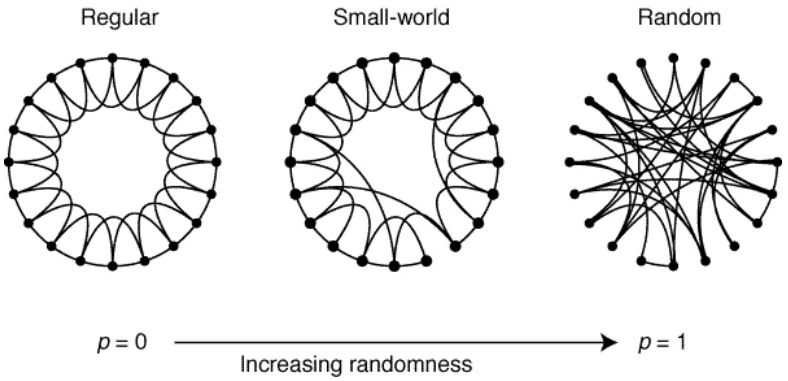


Figure 1: An illustration of the rewiring process of the Watts-Strogatz model from the original paper [24].

Example 3.15 (Random geometric graphs)**.** *Random geometric graphs* are another important class of network models. It has been shown that there are many real-world networks better modeled by random graphs than by any of the other models mentioned above [22, 18]. It has 3 parameters, and the parameter space is given by

$$\Theta = \mathbb{N} \times \mathbb{N} \times \left[0, \sqrt{d}\right] .$$

Given a triple $(N, d, r) \in \Theta$ we first sample for each node i in the network a point p_i uniformly at random in the d-dimensional euclidean cube $[0, 1]^d$. Next, we connect the nodes $\{i, j\}$ if the euclidean distance of their corresponding points is less than r. The probability of two arbitrary nodes being connected, $\mathbb{P}[X_{ij} = 1]$, can be analytically calculated for each d individually. Although a general formula for all d is not known yet [2], there are good estimates [1]. The classical approach would be to define $2d$ uniformly distributed independent random variables $U_1, V_1, \ldots U_d, V_d$ and subsequently define $D = \sqrt{(U_1 - V_1)^2 + \cdots + (U_d - V_d)^2}$. The sought probability is now given by $\mathbb{P}[X_{ij} = 1] = F_D(r)$, where F_D is the cumulative distribution function of the random variable D.
A closer examination of the model reveals that the random variables of the interaction probability are not fully independent, but depend on additional random variables which model the position of the nodes in euclidean space. We will only sketch how to formalize this issue as the general theory of this is not elementary. For each node i we can define a random vector X_i independently drawn from a uniform distribution on the cube $[0, 1]^d$. Intuitively the random variables X_{ij} become independent, in fact even deterministic, given that we know all the positions X_k; this is known as *conditional independence*.

Now, we did not define what it means to condition on a continuous random variable as all events of the form $[X_k = x]$ have probability zero. We will refer to the book by Fristedt [10] to address this issue. For further readings on random geometric graphs, we refer to the book by Mathew Penrose [21].

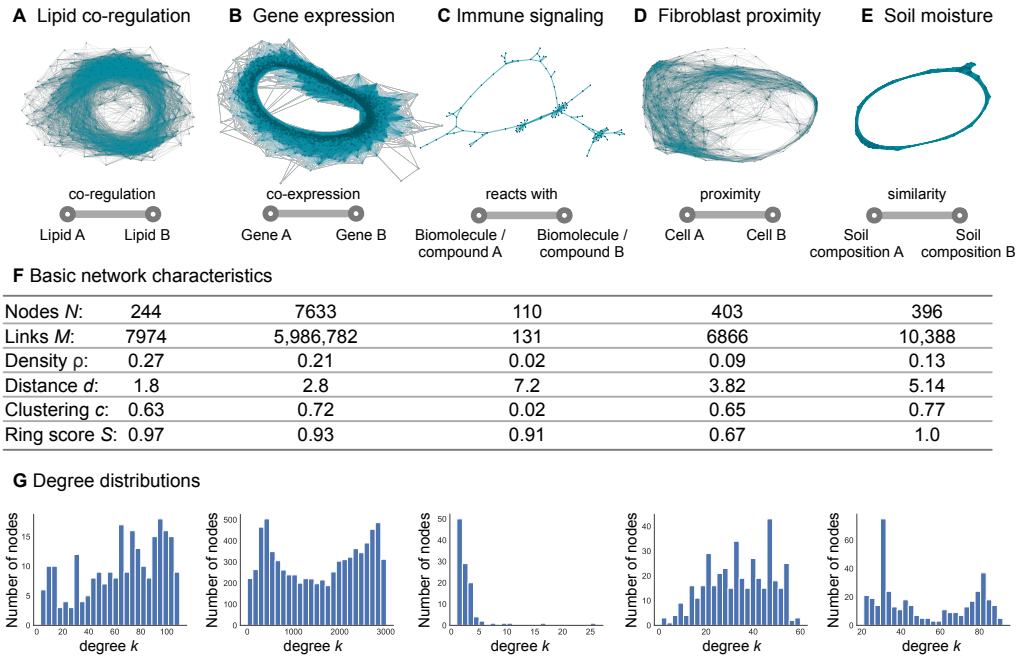
4 Scientific contributions in the paper

> Ring ring ring ring ring ring
> ring banana phone.
>
> Raffi and M. Creber,
> Bananaphone

A graph is an inherently abstract object in the sense that every visualization necessarily comes with an arbitrary decision. Formally, one would define an embedding into $\mathbb{R}^2$ or $\mathbb{R}^3$. Despite this fact, there exist embeddings of graphs into euclidean space that scientists believe can reveal valuable insights into the structure of the network. Algorithms belonging to the class of *force-directed graph drawing algorithms*, for example, assign forces among the set of edges and nodes to incorporate network-specific intuitions into the embedding. One particular instance, known as the*spring embedding*, utilizes Hooke's law to attract adjacent pairs of nodes towards each other, while simultaneously repulsive forces based on Coulomb's law are used to separate all pairs of nodes [15]. Instances of spring embeddings, using the Fruchterman–Reingold algorithm [12] to calculate local minima, can be seen in Fig. 2. Despite the vast structural diversity of the networks, a single realization of a spring embedding suggests a striking similarity between the networks in terms of a marked ring pattern in their global appearance. The above-defined network characteristics seen in Fig. 2F-G did not reveal any other common pattern: the networks range from sparsely to densely connected, from clustered to tree-like, exhibit both small and large average distances and diverse distributions of the number of links per node. No conventional network measure allowed us to pinpoint their global ring structure. Based on that observation we developed a "ring score" aimed to formalize this visual observation. As can be seen in the last row of Fig. 2F all depicted networks obtain a high ring-score deeming them similar in their ring-structure. To further examine the underlying mechanisms that lead to the emergence of ringiness we proposed a novel network model on 5 parameters.

4.1 Ring score

The idea of the ring score is based on the fact that the rank of the first homology group of a ring-shaped object is 1. The viewpoint we are taking is that for any given persistence diagram the most persistent homology class is regarded as a "true signal" whereas all other classes are seen as "noise." The more persistent the noise is, the less likely we believe that the underlying



Nodes *N*:	244	7633	110	403	396
Links *M*:	7974	5,986,782	131	6866	10,388
Density ρ:	0.27	0.21	0.02	0.09	0.13
Distance *d*:	1.8	2.8	7.2	3.82	5.14
Clustering *c*:	0.63	0.72	0.02	0.65	0.77
Ring score *S*:	0.97	0.93	0.91	0.67	1.0

Figure 2: Ring networks with a great variety of structural diversity.

structure is a ring. The question still remains of how to assign a persistence diagram to any given network. It turns out that using the Vietoris-Rips construction with the shortest path length distance is not sufficient enough to "detect" a ring-structure in networks we believe are ring-like. That is why we used the current flow distance as a more fine-grained distance measure. In summary, given a network $G = (V, E)$, the final ring score can be obtained in the following way:

1. Calculate the matrix of pairwise current flow distances D_{flow}.
2. Calculate the first persistence diagram $\text{dgm}_1(D_{\text{flow}})$ using the Vietoris-Rips complex.
3. Calculate the persistence sequence $\mathbf{p} = (p_0, p_1.p_2 \dots)$ of $\text{dgm}_1(D_{\text{flow}})$.
4. The score is now given by

$$\text{score}(G) = 1 - \sum_{i=1}^{\infty} \frac{p_i}{2^i}\,.$$

This ring score allowed for a rich and thorough analysis of a variety of real-world and synthetic networks. For the interested reader we refer to the paper.

4.2 A novel network model

Five principles

To gain a deeper understanding of ring structures in networks, as well as their structural and dynamic implications, we aimed to identify key principles that lead to their emergence. We established five principles that are shared among the diverse systems in Figure 2 and that collectively give rise to the emergence of ring structures in their network representation

1. The system is composed of individual elements equipped with an internal state.
2. The elements' internal states are coupled to an external signal.
3. The external signal displays hysteresis as it returns to a certain state.
4. The different elements exhibit diverse response profiles.
5. Two elements are more likely to be linked, if their response profiles are more similar.

General setup

Motivated by these principles we defined a procedure to generate a network model based on the following set of data

1. A number of elements N.
2. An internal state-space σ_{int}.
3. An external state-space σ_{ext}.
4. A set of functions $\mathcal{R} \subseteq \{r \colon \sigma_{\text{ext}} \to \sigma_{\text{int}}\}$ referred to as *response profiles*.
5. A probability measure $\mathbb{P}_{\mathcal{R}^N}$ together with a suitable σ-algebra on $\mathcal{R}^N$.
6. A is symmetric function $p : \mathcal{R} \times \mathcal{R} \to [0, 1]$ referred to as *interaction probability function*.

Since the only role of the state-spaces is to define the set of response profile, we encourage the reader to think of this construction as the quadruple $(N, \mathcal{R}, \mathbb{P}_{\mathcal{R}^N}, p)$ rather than the full sextuple $(N, \sigma_{\text{int}}, \sigma_{\text{ext}}, \mathcal{R}, \mathbb{P}_{\mathcal{R}^N}, p)$. Less formally, we have the number of elements in the system, N, a description of how a response profile can look like, R, a distribution describing how the response profiles are chosen, P, and a function describing how these response

profiles interact with each other, p. Note that by selecting an interaction probability function p we implicitly acquire a notion of similarity of two response profiles. We will discuss this matter in more detail below.

Network construction

Given a quadruple $(N, \mathcal{R}, \mathbb{P}_{\mathcal{R}^N}, p)$ we can perform the following network construction:

- Draw a vector of response profiles $\mathbf{r} = (r_i)_{i=0}^N \in \mathcal{R}^N$ according to a joint distribution given by $\mathbb{P}_{\mathcal{R}^N}$, meaning that the probability of $\mathbf{r}$ being in a set $A \subset \mathcal{R}^N$ is given by $\mathbb{P}_{\mathcal{R}^N}[A]$.

- Construct a network on N nodes such that each node i is associated with the response profile r_i defined in the previous step. Define further a link for each pair of nodes $\{i, j\}$ with a probability equal to the corresponding interaction probability $p(r_i, r_j)$. Formally, we define a single probability measure $\mathbb{P}_{\Omega_N}$ on the set of all graphs Ω_N that is uniquely defined by the interaction probability matrix that is given by the (conditionally) independent random variables X_{ij} such that

$$\mathbb{P}_{\Omega_N}[X_{ij} = 1] = p(r_i, r_j)\,.$$

  Note, that the X_{ij} are not by themselves independent, but only once we know $\mathbf{r}$, which itself is randomly distributed according to $\mathbb{P}_{\mathcal{R}^N}$.

In summary, we get a probability measure on $(\Omega_N, \mathfrak{P}(\Omega_N), \mathbb{P}_{\Omega_N})$ for each quadruple $(N, \mathcal{R}, \mathbb{P}_{\mathcal{R}^N}, p)$. If the quadruple itself is parametrized by a parameter space Θ, we obtain a network model on that exact parameter space.

Examples

Before we define our concrete implementation of the five principles within this class of networks, let us give some examples to build up intuition.

Example 4.1 (Erdős–Rényi model)**.** Let $(\sigma_{\text{int}}, \sigma_{\text{ext}}, \mathcal{R}, \mathbb{P}_{\mathcal{R}^N})$ be arbitrarily defined and let the interaction probability function be constant, i.e. $p_\varrho(r, r') = \varrho$ for all $r, r' \in \mathcal{R}$ and $\varrho \in [0, 1]$. Then the network model $(N, \mathcal{R}, \mathbb{P}_{\mathcal{R}^N}, p_\varrho)$ recapitulates the Erdős–Rényi model with parameters N and ϱ.

Example 4.2 (Regular ring lattice)**.** Let $\sigma_{\text{int}} = \{0, 1\}$ be a 2-state space, $\sigma_{\text{ext}} = [-\pi, \pi]/\{-\pi, \pi\}$ the circle and $\mathcal{R} = \{\delta_\theta : \sigma_{\text{ext}} \to \sigma_{\text{int}} \mid \theta \in \sigma_{\text{ext}}\} \cong \sigma_{\text{ext}}$

the set of all delta functions, where a delta function centered at θ is defined as

$$\delta_\theta(x) = \begin{cases} 1, & \text{if } x = \theta \\ 0, & \text{else.} \end{cases}$$

Define further the geodesic distance on the circle to be

$$d_{S^1}(\theta, \theta') = \min\left\{|\theta - \theta'|, 2\pi - |\theta - \theta'|\right\} ,$$

and subsequently the interaction probability function to be

$$p_\varrho(\delta_\theta, \delta_{\theta'}) = \begin{cases} 1, & \text{if } d_{S^1}(\theta, \theta') \leq \varrho\pi \\ 0, & \text{else} , \end{cases}$$

with $\varrho \in [0, 1]$. Let further $\mathbb{P}_{\mathcal{R}^N}$ be concentrated on the single outcome, in which the positions of the delta functions are equidistantly places along the circle σ_{ext}, i.e. if

$$\mathbb{P}[\{\mathbf{r}\}] = 1 \quad \text{for} \quad \mathbf{r} = \left(\frac{2\pi i}{N}\right)_{i=1}^{N} .$$

Then the network model $(N, \mathcal{R}, \mathbb{P}_{\mathcal{R}^N}, p_\varrho)$ recapitulates the ring lattice from the Watts-Strogatz model with paramters N and k if we identify the parameter for the average degree via $k = \varrho(N - 1)$.

In case each response profile is drawn independently from the same distribution we can conceptionally simplify the setup by letting the probability measure be defined on $\mathcal{R}$ instead of $\mathcal{R}^N$. Denoting this new probability measure by $\mathbb{P}$, the joint distribution can be recovered via

$$\mathbb{P}_{\mathcal{R}^N}[A] = \prod_{i=1}^{N} \mathbb{P}_{\mathcal{R}}[A_i] ,$$

for each "box" $A = (A_i)_{i=1}^N \subset \mathcal{R}^N$ with $A_i \subset \mathcal{R}$.

Example 4.3 (Random geometric graph)**.** Let $\sigma_{\text{int}} = \{0, 1\}$ be a 2-state space, $\sigma_{\text{ext}} = [0, 1]^d$ the d-dimensional cube and $\mathcal{R}_d = \{\delta_{\mathbf{x}} : \sigma_{\text{ext}} \to \sigma_{\text{int}} \mid \mathbf{x} \in \sigma_{\text{ext}}\} \cong \sigma_{\text{ext}}$ the set of all delta functions as defined above. Let further d_2 be the euclidean distance on σ_{ext} and define susquently the interaction probability function to be

$$p_r(\delta_{\mathbf{x}}, \delta_{\mathbf{y}}) = \begin{cases} 1, & \text{if } d_2(\mathbf{x}, \mathbf{y}) \leq r \\ 0, & \text{else} , \end{cases}$$

with $r \in \left[0, \sqrt{d}\right]$. If now $\mathbb{P}_{\mathcal{R}}$ is the uniform distribution on $\mathcal{R}_d \cong [0, 1]^d$, then the network model $(N, \mathcal{R}_d, \mathbb{P}_{\mathcal{R}_d}, p_r)$ recapitulates the random geometric graph model with parameters N, d and r.

Interaction probability function

The fifth principle assumes a notion of similarity for the response profiles, but in the generic setting above, this notion is only implicitly given by the interaction probability function $p : \mathcal{R} \times \mathcal{R} \to [0, 1]$. To fully understand the mechanisms giving rise to specific network features, our idea was to factor this function through other functions that are more easily interpretable. To that end, we start with the arguably most familiar concept in mathematics related to the idea of similarity; namely a metric, d, as a measure of dissimilarity. Next, we normalize and reverse that metric via a function s to get a notion of similarity in $[0, 1]$ As a final step, we can control further desired network properties by converting this similarity via a function F instead of interpreting the outcome of s directly as interaction probability. In summary, we get the following commuting diagram

$$\begin{array}{ccc} \mathbb{R}_{\geq 0} & \xrightarrow{s} & [0,1] \\ {\scriptstyle d}\uparrow & & \downarrow{\scriptstyle F} \\ \mathcal{R} \times \mathcal{R} & \xrightarrow{p} & [0,1] \end{array} .$$

Assuming the metric d on $\mathcal{R}$ is given, there are a few mild assumptions on s and F to make this factorization of p meaningful:

- The function $s : \mathbb{R}_{\geq 0} \to [0, 1]$ is required to to be monotonously decreasing.
- The function $F : [0, 1] \to [0, 1]$ is required to be monotonously increasing. As an additional, technical condition we require it to be right-continuous, guaranteeing it to be a cumulative distribution function of some probability distribution on $[0, 1]$.

The fact that $F = F_B$ can be interpreted as a cumulative distribution function of a random variable B allows us to translate probabilistic properties of B into expected network properties. As an example, the mean of B correlates inversely to the expected network density: the higher the mean, the lower the network density. Another rule of thumb is the role of the variance of B corresponding to the influence of the similarity, and hence the structure of $\mathcal{R}$, to the overall interaction probability: the lower the variance, the more regular the network becomes. This leads to the unique role of two families of distributions:

Delta: As the distribution with minimal variance, taking F to be the cumulative distribution function of the delta distribution results

in a noise-free translation of the similarity structure of $\mathcal{R}$ to the network structure. A closer examination of the examples above reveals that the regular ring lattice, as well as the random geometric graph, can be expressed in that way.

Bernoulli: According to the Bhatia–Davis inequality, any distribution B on $[0,1]$ with mean μ has variance at most $(1-\mu)\mu$ and equality holds if and only if B is Bernoulli distributed [4]. Indeed, the cumulative distribution function, F_B, is (almost) constant reducing (almost) any model with this choice of $F = F_B$ to the case of an Erdős–Rényi model with $\varrho = 1 - \mu$, as discussed in the example above.

Our network model

In the paper, we presented a model to get a better insight into the mechanisms that lead to the emergence of ring-structures. To that end, we defined a network model $(N, \mathcal{R}_a, \mathbb{P}_\kappa, p_{a,\mu,\nu})$ on the parameters N, a, κ, μ, η as follows:

1. The number of nodes is given by N.
2. The internal state-space is given by a *two-state space* $\sigma_{\text{int}} = \{0, 1\}$.
3. The external state-space is given by a *circle* $\sigma_{\text{ext}} = [-\pi, \pi]/\{-\pi, \pi\}$.
4. The response profiles are given by *box functions* of length $2\pi a$, i.e. $\mathcal{R}_a = \left\{ \mathbb{1}_{[\theta-\pi a, \theta+\pi a]} \mid \theta \in \sigma_{\text{ext}} \right\} \cong \sigma_{\text{ext}}$.
5. The probability distribution on $\mathcal{R}_a$, denoted by $\mathbb{P}_\kappa$, is given by the *von Mises distribution* on the circle with concentration parameter κ. This means, that the density function of this distribution is given by

$$f_{\mathbb{P}_\kappa} = \frac{e^{\kappa \cos\theta}}{2\pi I_0(\kappa)} ,$$

   where $I_0(\kappa)$ is a normalization function known as the *modified Bessel function* of order 0.
6. The interaction probability function is given by $p_{a,\mu,\eta} = F_{\mu,\eta} \circ s_a \circ d$, where
   - the metric d is given by the *geodesic distance* on the circle

$$d(\theta, \theta') = \min\left\{ |\theta - \theta'|, 2\pi - |\theta - \theta'| \right\} , \tag{d}$$

- the similarity function is given by the *normalized overlap* of two box functions

$$s_a(d) = \max\left\{0, 1 - \frac{d}{2\pi a}\right\}, \tag{s}$$

- the function F is given by the cumulative distribution function of a *Beta distribution* with mean μ and variance $\eta \cdot (1-\mu)\mu$. We can write it down explicitly, as

$$F_{\mu,\eta}(s) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^s x^{\alpha-1}(1-x)^{\beta-1} dx\,, \tag{F}$$

  with $\alpha = \mu\frac{1-\nu}{\nu}$, $\beta = (1-\mu)\frac{1-\mu}{\mu}$ and Γ being the *Gamma function*, the commonly used extension of the factorial function.

By changing the length of the box functions, the parameter a regulates the "activity window" of all elements in the system. The parameter κ models the strength of preference of the elements towards a specific external signal. The overall interaction strength can be regulated by the parameter μ. This corresponds directly to the expected network density. For this reason, we replaced this parameter with the expected density ϱ in the model presented in the paper. We emphasize, however, that this correspondence is highly non-trivial and, so far, we only managed to get analytically closed forms for specific parameter-combinations. Lastly, the parameter ν models the amount of randomness, $\eta = 0$ being a highly regular ring structure and $\eta = 1$ resulting in the Erdős–Rényi model.

References

[1] RS Anderssen, RP Brent, DJ Daley, and PAP Moran. Concerning \int_0^1\int_0^1(x_1^2++x_k^2)^1/2dx_1,dx_k and a taylor series method. *SIAM Journal on Applied Mathematics*, 30(1):22–30, 1976.

[2] David H Bailey, Jonathan M Borwein, and Richard E Crandall. Box integrals. *Journal of Computational and Applied Mathematics*, 206(1):196–208, 2007.

[3] A L Barabási and R Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[4] Rajendra Bhatia and Chandler Davis. A better bound on the variance. *The American Mathematical Monthly*, 107(4):353–357, 2000.

[5] Nicolas Bourbaki. *Topologie algébrique: Chapitres 1 à 4*. Springer, 2016.

[6] Reinhard Diestel. Graph theory, volume 173 of. *Graduate texts in mathematics*, page 7, 2012.

[7] H Edelsbrunner, D Letscher, and A Zomorodian. Topological persistence and simplification. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 454–463. ieeexplore.ieee.org, November 2000.

[8] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[9] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[10] Bert E Fristedt and Lawrence F Gray. *A modern approach to probability theory*. Springer Science & Business Media, 2013.

[11] Patrizio Frosini and Claudia Landi. Size theory as a topological tool for computer vision. *Pattern Recognition and Image Analysis*, 9(4):596–603, 1999.

[12] Thomas MJ Fruchterman and Edward M Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

[13] E N Gilbert. Random graphs. *Ann. Math. Stat.*, 30(4):1141–1144, 1959.

[14] Achim Klenke. *Probability theory: a comprehensive course*. Springer Science & Business Media, 2013.

[15] Stephen G Kobourov. Spring embedders and force directed graph drawing algorithms. *arXiv preprint arXiv:1201.3011*, 2012.

[16] R Duncan Luce and Albert D Perry. A method of matrix analysis of group structure. *Psychometrika*, 14(2):95–116, 1949.

[17] Peter McCullagh. What is a statistical model? *Annals of statistics*, pages 1225–1267, 2002.

[18] Maziar Nekovee. Worm epidemics in wireless ad hoc networks. *New Journal of Physics*, 9(6):189, 2007.

[19] M E J Newman. A measure of betweenness centrality based on random walks. *Soc. Networks*, 27(1):39–54, January 2005.

[20] Mark Newman. *Networks*. Oxford university press, 2018.

[21] Mathew Penrose et al. *Random geometric graphs*, volume 5. Oxford university press, 2003.

[22] Natasa Pržulj, Derek G Corneil, and Igor Jurisica. Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18):3508–3515, 2004.

[23] Vanessa Robins. Towards computing homology from finite approximations. In *Topology proceedings*, volume 24, pages 503–532, 1999.

[24] D J Watts and S H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.