

Metric Structures on Networks and Applications

Samir Chowdhury¹ and Facundo Mémoli²

Abstract—It is a common belief that biological networks are determined by their network motifs. Empirical results show that simulated networks with similar motif structure also reflect similarity across other structural measures. In this work we tackle this question from a theoretical perspective and announce our results proposing a family of invariants, the *motif sets*, as a natural modelization of “bags of motifs”. We prove that these motif sets precisely characterize networks which are *weakly isomorphic*, in the sense that the networks are at 0-distance with respect to a certain pseudometric on the space of all networks. This notion differs from the usual notion of (strong) isomorphism. We further explore a notion of *skeleton* of a network and establish that this construction mediates between strong and weak isomorphism, in the sense that two networks are weakly isomorphic if and only if there exists a strong isomorphism between their respective skeleta.

I. INTRODUCTION

Networks which show the relationships within and between complex systems are key tools in a variety of current scientific areas. In the domain of bioinformatics, networks have been used to represent molecular activity [12], metabolic pathways [10], functional relations between enzyme clusters [8], genetic regulation [13], and brain organization [11], [9]. Other examples of networks include social networks [3], information networks such as the World Wide Web [5], and technological networks such as the electric power grid [15]. For a more comprehensive overview of the literature on complex networks, consult [7].

One of the prevalent hypotheses used in systems biology and network analysis is that complex networks are assembled from simpler subnetworks called *motifs* [13], [14], [2], [6], [1]. Motifs have been used to characterize transcription regulation networks and protein-protein interaction [16]. Sporns and Kötter have used motifs to simulate network datasets that resemble real brain networks across a variety of structural measures [14]. These considerations motivate the following theoretical question:

Is it possible to reconstruct, up to isomorphism, a network from the knowledge of its subnetworks?

The goal of this work is to provide an answer to the question above. In order to be able to reason about and eventually answer this question one needs to define several concepts, for example: what is the definition of network that one should use, what is a suitable notion of isomorphism

between two networks, and what is meant by “knowledge of its subnetworks”?

Biological networks may have asymmetric edge relations and autoregulatory properties, so from a physiological perspective, they should be represented by directed, weighted networks with self-loops, without necessarily imposing any additional conditions on edge weights. A natural choice is therefore to represent networks as weighted directed graphs, where the weights are allowed to be real numbers. One natural notion of isomorphism in this setting is the usual notion of graph isomorphism which in our context we call *strong isomorphism*. With regards to subnetworks: we organize all the motifs present in a given network X into *motif sets*: these “bags”, denoted $M_n(X)$ for $n = 1, 2, \dots$, contain all motifs involving at most n vertices. For example, $M_3(X)$ contains all possible motifs of size 3 or smaller and so on. A precise definition is given §II.

Contributions.

In this work we first verify that under the notion of strong isomorphism the question above cannot admit a positive answer. As a second step, we identify an alternative notion, *weak isomorphism*, under which we establish an affirmative answer to the question. As a third development, we study and clarify the precise relationship between weak and strong isomorphism: we introduce the *skeleton* of a network X as a minimalistic representation of the structural information contained in X and prove that *two networks are weakly isomorphic if and only if their skeleta are strongly isomorphic*.

In order to be able to formulate our answer to the network reconstruction question posed above we utilize a notion of distance on the space of all possible networks with the property that two networks are weakly isomorphic exactly when this distance vanishes. As a by-product of this construction we identify a number of *network signatures* or *invariants* that are *stable* under perturbation in the sense of this distance. Invariants are useful because they can be computed quickly and can be used to distinguish between distinct networks. Lower bounds for the distance between two networks arise by computing distances between the (often simpler) invariants associated to the networks.

In this paper we briefly describe our framework and result — a complete presentation is forthcoming.

II. NETWORKS, ISOMORPHISM, AND MOTIFS

Let X be a finite set, and let ω_X be a function from $X \times X$ to \mathbb{R} . By a network, we will mean a pair (X, ω_X) . We will denote the universe of all networks by \mathcal{N} . Notice that we do not make any assumptions on the range of ω_X : in particular,

*This work was not supported by NSF grant IIS-1422400

¹Samir Chowdhury is a Ph.D. student with the Department of Mathematics, The Ohio State University, Columbus, OH 43210 chowdhury.57@osu.edu

²Facundo Mémoli is with the Departments of Mathematics and Computer Science and Engineering, The Ohio State University, Columbus, OH 43210, USA memoli@math.osu.edu

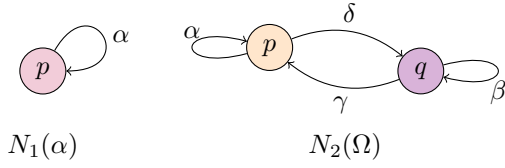


Fig. 1. Networks over one and two nodes with their weight functions.

ω_X could be negative. Our framework therefore includes but is not limited to finite metric spaces. In particular, our networks are *directed*.

We will refer to the points of X as *nodes* and ω_X as the *weight function* of X . Pairs of nodes will be called *edges*. The information contained in a network should be preserved when we relabel the nodes in a compatible way; we formalize this idea by the following notion of *strong isomorphism* of networks: To say (X, ω_X) and (Y, ω_Y) are *strongly isomorphic* means that there exists a bijection $\varphi : X \rightarrow Y$ such that $\omega_X(x, x') = \omega_Y(\varphi(x), \varphi(x'))$ for all $x, x' \in X$. Note that this definition is exactly the usual notion of isomorphism between weighted, directed graphs. We will denote a strong isomorphism between networks by $X \cong^s Y$. This definition captures the simple idea that permuting the nodes should not affect the structure of the network. Notice that a *necessary* condition for two networks to be strongly isomorphic is that they have the *same* number of nodes.

The following examples illustrate these concepts.

Example 1: Networks with one or two nodes will be very instructive in providing examples and counterexamples, so we introduce them now with some special terminology. See Figure 1.

- Networks with one node: any network with exactly one node p can be specified by a real number α , and we will often denote them by $N_1(\alpha)$. From our definition, it follows that networks $N_1(\alpha)$ and $N_1(\alpha')$ are strongly isomorphic if and only if $\alpha = \alpha'$.
- Given α, β, δ and γ in \mathbb{R} let $\Omega = \begin{pmatrix} \alpha & \delta \\ \gamma & \beta \end{pmatrix}$. This matrix induces a network $N_2(\Omega)$ over two nodes. Any network with two nodes can be represented this way. Notice that $N_1(\alpha) \not\cong^s N_2(\Omega)$ since they have different numbers of nodes. Given $\Omega, \Omega' \in \mathbb{R}^{2 \times 2}$, $N_2(\Omega) \cong^s N_2(\Omega') \Leftrightarrow$ there exists a permutation matrix P of size 2×2 such that $\Omega' = P \Omega P^T$.
- Any k -by- k matrix $\Sigma \in \mathbb{R}^{k \times k}$ induces a network on k nodes, which we refer to as $N_k(\Sigma)$. Notice that $N_k(\Sigma) \cong N_\ell(\Sigma')$ if and only if $k = \ell$ and there exists a permutation matrix P of size k such that $\Sigma' = P \Sigma P^T$.

In what follows we will precisely model a motif on a network X as any subnetwork of X together with its inherited weights. For a sequence $(x_i)_{i=1}^n$ of nodes in a network X , we will denote the associated weight matrix by $((\omega_X(x_i, x_j)))_{i,j=1}^n$: this weight matrix contains all the information there is, up to isomorphism, about the motif defined by the points $(x_i)_{i=1}^n$.

Definition 1 (motif set): For each $n \in \mathbb{N}$ and each $X \in \mathcal{N}$,

define $\Psi_X^n : X^n \rightarrow \mathbb{R}^{n \times n}$ to be the map $(x_1, \dots, x_n) \mapsto ((\omega_X(x_i, x_j)))_{i,j=1}^n$. Note that Ψ_X^n is simply a map that sends each sequence to its corresponding weight matrix. Let $F(\mathbb{R}^{n \times n})$ denote the finite subsets of $\mathbb{R}^{n \times n}$. Then let $M_n : \mathcal{N} \rightarrow F(\mathbb{R}^{n \times n})$ denote the map defined by

$$(X, \omega_X) \mapsto \{\Psi_X^n(x_1, \dots, x_n) : x_1, \dots, x_n \in X\}.$$

We refer to $M_n(X)$ as the n -motif set for X .

Notice that for a fixed natural number n , the set $M_n(X)$ is a *finite* subset of $\mathbb{R}^{n \times n}$. The interpretation is that $M_n(X)$ is a bag containing all the motifs of X that one can form by looking at all subnetworks of size n (with repetitions).

Example 2: Consider the two networks from Example 1. Then we have $M_1(N_2(\Omega)) = \{\alpha, \beta\}$ and

$$M_2(N_2(\Omega)) = \left\{ \begin{pmatrix} \alpha & \alpha \\ \alpha & \alpha \end{pmatrix}, \begin{pmatrix} \beta & \beta \\ \beta & \beta \end{pmatrix}, \begin{pmatrix} \alpha & \delta \\ \gamma & \beta \end{pmatrix}, \begin{pmatrix} \beta & \gamma \\ \delta & \alpha \end{pmatrix} \right\},$$

$$M_2(N_1(\alpha)) = \left\{ \begin{pmatrix} \alpha & \alpha \\ \alpha & \alpha \end{pmatrix} \right\}.$$

In line with our discussion in the introduction, we wish to examine the extent to which motif sets determine the structure of a network. With the minimal definitions we have so far, we now formulate the conjecture that strong isomorphism of networks is characterized by equality of motif sets.

Conjecture 1 (strong isomorphism conjecture): Let $X, Y \in \mathcal{N}$. Then $M_n(X) = M_n(Y)$ for all $n \in \mathbb{N}$ if and only if $X \cong^s Y$.

This conjecture is in general *false*: indeed, $M_n(N_1(1)) = M_n(N_2(1_{2 \times 2})) = \{1_{n \times n}\}$ for all $n \in \mathbb{N}$, but we know the two intervening networks are not strongly isomorphic (since they have different cardinalities). Here $1_{n \times n}$ denotes the all-ones matrix of size $n \times n$.

It turns out, surprisingly, that a weaker but very natural version of the conjecture above holds in full generality. We first declare that networks X and Y are *weakly isomorphic*, denoted $X \cong^w Y$, if there exists a finite set Z and *surjective* maps $\phi_X : Z \rightarrow X$ and $\phi_Y : Z \rightarrow Y$ such that

$$\omega_X(\phi_X(z), \phi_X(z')) = \omega_Y(\phi_Y(z), \phi_Y(z')) \text{ for all } z, z' \in Z.$$

Notice that this is in fact a *relaxation of the notion of strong isomorphism*: indeed, if in addition to being surjective, we require the maps ϕ_X and ϕ_Y to be injective, then these maps are forced to be *bijective* so that the strong notion of isomorphism is recovered: in this case the map $\phi_Y \circ \phi_X^{-1} : X \rightarrow Y$ would be a weight preserving bijection between the networks X and Y .

For example, networks $X = N_1(1)$ and $Y = N_2(1_{2 \times 2})$, with node sets $\{p\}$ and $\{q, r\}$ respectively, are weakly (but not strongly) isomorphic: let $Z = \{q, r\}$, let ϕ_Y be the identity map on $\{q, r\}$ and let ϕ_X be the map that sends q, r to p . Since all the weights are equal to 1, both maps satisfy the necessary conditions on the weights.

Notation. For any set X , we will denote its cardinality by $\text{card}(X)$.

Having relaxed the notion of isomorphism we formulate a less stringent version of Conjecture 1: namely we conjecture that *weak* isomorphism is completely characterized by equality of motif sets.

Conjecture 2 (weak isomorphism conjecture): Let $X, Y \in \mathcal{N}$. Then $M_n(X) = M_n(Y)$ for all $n \in \mathbb{N}$ if and only if $X \cong^w Y$.

In sharp contrast with Conjecture 1, this conjecture is true.

In order to elicit a connection between the concept of structural network equivalence — as provided by the notion of strong isomorphism — and the notion of motifs we weakened the initial notion of (strong) isomorphism to uncover a precise connection with the equivalence of motif sets. Of course strong isomorphism implies weak isomorphism, and weak isomorphism does not in general imply strong isomorphism, as we saw in the simple examples discussed above. One may nevertheless wonder whether strong and weak isomorphism may be related in the following sense:

Could it be that weak isomorphism between two networks X and Y might imply that some essential network structures associated to X and Y are strongly isomorphic?

It turns that the answer to this question is positive:

Theorem 3: Let $(X, \omega_X), (Y, \omega_Y) \in \mathcal{N}$. Then $X \cong^w Y$ if and only if $\text{skel}(X) \cong^s \text{skel}(Y)$.

Here $\text{skel}(X)$ and $\text{skel}(Y)$ are called the *skeleta* of X and Y , respectively. The *skeleton* $\text{skel}(X)$ of a network X will turn out to be the *smallest* network that is weakly isomorphic to X , where smallest means that its cardinality is minimal. Thus, if two networks are weakly isomorphic, then one is essentially a “fattening” of the other.

Comments.

In order to present a proof of Conjecture 2 we need to introduce some tools that help in organizing and decoding the structure of space \mathcal{N} of all networks with respect to isomorphism. Intuitively, we are trying to understand whether certain invariants, the motif sets, are able to discriminate networks up to weak isomorphism.

The problem of discriminating between networks based on the *binary* notion of whether they are exactly weakly isomorphic or not is rather rigid. To tackle the conjecture we relax this situation by indirectly introducing a notion of “almost isomorphism”. This is achieved by considering a notion of *distance between networks* which by construction will be compatible with weak isomorphism: it will turn out that two networks will be at zero distance if and only if they are weakly isomorphic. Introducing this distance between networks not only provides us with a path for proving Conjecture 2 but also this notion of distance will be crucial in providing an intuition for what weakly isomorphic networks “look” like, clarifying the notion of skeleton mentioned above.

As by-product with practical implications, we will use this distance to project partial information about a network into the real line, where comparison is almost trivial. This has potential applications in the classification of network data.

Finally, we note that apart from weakening the notion of isomorphism, a natural choice for trying to recover a true statement from Conjecture 1 is to strengthen the assumptions

on the networks. Under the additional assumption that the edge weights are unique, the strong conjecture is true.

Remark 4: Our definition of motif sets is inspired by a definition made by Gromov, termed “curvature classes” in the context of compact metric spaces [4, Chapter 3].

III. DISCUSSION

As a by-product of our constructions we identify a number of computable lower bounds for the distance between networks, several of which have potential applications in network comparison. Full details will be given in an upcoming publication.

REFERENCES

- [1] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2006.
- [2] Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, 2007.
- [3] Easley David and Kleinberg Jon. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, New York, NY, USA, 2010.
- [4] Misha Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Birkhäuser Boston Inc., 1999.
- [5] Jon M Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew S Tomkins. The web as a graph: measurements, models, and methods. In *Computing and combinatorics*, pages 1–17. Springer, 1999.
- [6] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
- [7] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [8] Hiroyuki Ogata, Wataru Fujibuchi, Susumu Goto, and Minoru Kanehisa. A heuristic graph comparison algorithm and its application to detect functionally related enzyme clusters. *Nucleic acids research*, 28(20):4021–4028, 2000.
- [9] Luiz Pessoa. Understanding brain networks and brain organization. *Physics of life reviews*, 11(3):400–435, 2014.
- [10] Ron Y Pinter, Oleg Rokhlenko, Esti Yeger-Lotem, and Michal Ziv-Ukelson. Alignment of metabolic pathways. *Bioinformatics*, 21(16):3401–3408, 2005.
- [11] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.
- [12] Roded Sharan and Trey Ideker. Modeling cellular machinery through biological network comparison. *Nature biotechnology*, 24(4):427–433, 2006.
- [13] Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of *escherichia coli*. *Nature genetics*, 31(1):64–68, 2002.
- [14] Olaf Sporns and Rolf Kötter. Motifs in brain networks. *PLoS biology*, 2(11):e369, 2004.
- [15] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-world networks. *nature*, 393(6684):440–442, 1998.
- [16] Esti Yeger-Lotem, Shmuel Sattath, Nadav Kashtan, Shalev Itzkovitz, Ron Milo, Ron Y Pinter, Uri Alon, and Hanah Margalit. Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):5934–5939, 2004.