Counting Simplicial Pairs in Hypergraphs

Jordan Barrett[1], Paweł Prałat[1] Aaron Smith[2], and François Théberge[3]

[1] Department of Mathematics, Toronto Metropolitan University, Toronto, ON, Canada, {jordan.barrett,pralat}@torontomu.ca
[2] Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada, asmi28@uOttawa.ca
[3] Tutte Institute for Mathematics and Computing, Ottawa, ON, Canada, theberge@ieee.org

Abstract. We present two ways to measure the simplicial nature of a hypergraph: the simplicial ratio and the simplicial matrix. We show that the simplicial ratio captures the frequency, as well as the rarity, of simplicial interactions in a hypergraph while the simplicial matrix provides more fine-grained details. We then compute the simplicial ratio, as well as the simplicial matrix, for 10 real-world hypergraphs and, from the data collected, hypothesize that simplicial interactions are more and more *deliberate* as edge size increases. We then present a new Chung-Lu model that includes a parameter controlling (in expectation) the frequency of simplicial interactions. We use this new model, as well as the real-world hypergraphs, to show that multiple stochastic processes exhibit different behaviour when performed on simplicial hypergraphs vs. non-simplicial hypergraphs.

Keywords: simplicial pairs, hypergraphs

1 Introduction

Many datasets that are typically represented as graphs would be more accurately represented as hypergraphs. For example, in the graph representation of a collaboration dataset, authors are represented as vertices and an edge exists between two vertices if the corresponding authors wrote a paper together [21]. Using this representation, it is impossible to distinguish between a three-author paper and three separate two-author papers. In contrast, when we represent a collaboration dataset as a hypergraph we can clearly distinguish between a three-author paper (a single hyperedge) and three separate two-author papers (three distinct hyperedges). Hypergraph representations have proven to be useful for studying collaboration datasets [8], protein complexes and metabolic reactions [6], and many other datasets that are traditionally represented as graphs [19]. Moreover, after many years of intense research using graph theory in modelling and mining complex networks [5, 7, 11, 20], hypergraph theory has started to gain considerable traction [1–4, 13, 10, 12]. It is becoming clear to both researchers and practitioners that higher-order representations are needed to study datasets involving higher-order interactions [3, 15, 23, 19].

Similar to hypergraph representations, simplicial complexes provide another way to represent datasets with higher-order interactions and, in some cases, it is not clear what the better model is for a given dataset [14, 24, 25]. The notion of *simpliciality* was first introduced by Landry, Young and Eikmeier in [17] as a way of describing how closely a hypergraph resembles its simplicial closure. In their work, they discover that many hypergraphs built from real-world data, although not actually simplicial complexes, resemble their simplicial closures more closely than random hypergraphs. In a similar but distinct study, LaRock and Lambiotte in [18] find that real-world hypergraphs often contain more instances of hyperedges contained in other hyperedges than in random hypergraphs. The results found in these two papers suggest that real-world hypergraphs are organized in a way where many of the small hyperedges live inside larger hyperedges. In our work, we pursue this idea further and define a ratio and a matrix for hypergraphs, which we call the *simplicial ratio* and *simplicial matrix* respectively, based on the number of instances of hyperedges inside other hyperedges compared to that of a null model.

1.1 Notation

For the duration of the paper, we use the terms graph and edge in lieu of hypergraph and hyperedge.

A graph G is a pair $(V(G), E(G))$ where $V(G)$ is a set of vertices and $E(G)$ is a collection of edges, i.e., a collection of subsets of vertices. We insist that $\emptyset \notin E(G)$ for any graph G. In general, for a graph G and edge $e \in E(G)$, it is acceptable that $|e| = 1$. In this paper, however, we forbid such edges and consider only edges of size at least 2. We write $[n] := \{1, \ldots, n\}$ and typically label the vertices in G as $[n]$. A subgraph of a graph G is any graph $H = (V(H), E(H))$ with $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ (note that, as H is itself a graph, any edge $e \in E(H)$ contains only vertices in $V(H)$). For $e \in E(G)$, write $|e|$ for the size of e and, for each positive integer k, define

$$E_k(G) := \{e \in E(G), |e| = k\}\,.$$

If $E_k(G) = E(G)$ for some $k > 0$, then we call G a k-uniform graph. Note that, for any graph G, the graph $G_k := (V(G), E_k(G))$ is a k-uniform subgraph of G, and

$$G = \bigcup_{k>0} G_k\,,$$

and thus every graph is the edge-disjoint union of uniform subgraphs.

1.2 Measures for simpliciality

In [17], Landry, Young and Eikmeier establish three distinct measures quantifying how close a graph is to a simplicial complex. The first measure they establish is the *simplicial fraction*. Given a graph G, let $S \subseteq E(G)$ be the set of edges

such that $e \in S$ if and only if $|e| \geq 3$ and, for all $f \subseteq e$ with $|f| \geq 2$, $f \in E(G)$. Then the *simplicial fraction* of G, written $\sigma_{\mathrm{SF}}(G)$, is defined as

$$\sigma_{\mathrm{SF}}(G) := \frac{|S|}{\left|\bigcup_{k\geq 3} E_k(G)\right|}\,.$$

In words, $\sigma_{\mathrm{SF}}(G)$ is the proportion of edges of size at least 3 in $E(G)$ that satisfy downward closure.

The second and third measures Landry, Young and Eikmeier establish are the *edit simpliciality* and the *face edit simpliciality*, respectively. For a graph G, define the *k-closure*, written $\overline{G}_k$, as the graph $(V(\overline{G}_k), E(\overline{G}_k))$ where

$$\begin{aligned} V(\overline{G}_k) &= V(G), \\ E(\overline{G}_k) &= \left\{ e \subseteq V(G) \;\middle|\; |e| \geq k \text{ and } e \subseteq f \text{ for some } f \in E(G) \right\}. \end{aligned}$$

Then the *edit simpliciality* of G, written $\sigma_{\mathrm{ES}}(G)$, is defined as

$$\sigma_{\mathrm{ES}}(G) := \frac{|E(G)|}{|E(\overline{G}_2)|}\,.$$

Thus, $1 - \sigma_{\mathrm{ES}}(G)$ is the (normalized) number of additional edges needed to turn G into its 2-closure. Similarly, the *face edit simpliciality* of G, written $\sigma_{\mathrm{FES}}(G)$, is the average edit simpliciality across all induced subgraphs defined by maximal edges (edges not contained in other edges) in $\bigcup_{k\geq 3} E_k(G)$.

Using the three measures defined above, Landry, Young and Eikmeier show that real-world graphs are significantly more simplicial than graphs sampled from random models. However, they also note some unique short-comings of each measure. In the following example, we show an additional short-coming that is shared among all three measures, namely, that none of the measures are good indicators of how common it is to see edges inside of other edges in the graph.

Example 1. Let G_1 and G_2 be as shown in Figure 1. There is a clear, strong simplicial structure in G_1, and there is clearly no simplicial structure in G_2. However, in both graphs, the simplicial fraction is 0 (none of the edges satisfy downward closure). Moreover, the edit simpliciality of G_1 is $4/57 \approx 0.07$ and of G_2 is $3/41 \approx 0.07$. Likewise, the face edit simpliciality of G_1 is $4/57 \approx 0.07$ and of G_2 is

$$\frac{1}{3}\left(\frac{1}{26} + \frac{1}{11} + \frac{1}{4}\right) \approx 0.13\,.$$

Thus, G_1 and G_2 are equally simplicial according to the simplicial fraction and edit simpliciality and, more strikingly, G_1 is *less* simplicial than G_2 according to the face edit simpliciality.

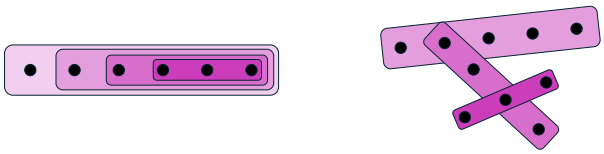

Fig. 1. (left) a graph G_1 with 6 vertices and 4 edges, and (right) a graph G_2 with 10 vertices and 3 edges. We have $\sigma_{\mathrm{SF}}(G_1) = 0$, $\sigma_{\mathrm{ES}}(G_1) \approx 0.07$, $\sigma_{\mathrm{FES}}(G_1) \approx 0.07$, and $\sigma_{\mathrm{SF}}(G_2) = 0$, $\sigma_{\mathrm{ES}}(G_2) \approx 0.07$, $\sigma_{\mathrm{FES}}(G_2) \approx 0.13$.

2 A new approach to simpliciality

We aim to quantify a graph based on the frequency and rarity of edges inside other edges when compared to a null model. Thoughout this section, as well as the remainder of the paper, we reference the Hypergraph Chung-Lu model and we write $\hat{G} \sim \mathrm{CL}(G)$ to mean $\hat{G}$ is sampled as a Chung-Lu model based on G. We give an algorithm for building this model, conditioned on the number of edges, in the journal version of the paper, and point the reader to [9] for a full description of the model.

2.1 The simplicial ratio

For a graph G, a *simplicial pair in* G is a pair of distinct edges $e_1, e_2 \in E(G)$ with $e_1 \subset e_2$. Let $\mathrm{sp}(G)$ be the number of simplicial pairs in G.

Let G be a graph and let $\hat{G} \sim \mathrm{CL}(G)$ conditioned on $\hat{G}$ having no multiset edges. Then the *simplicial ratio*, denoted by $\sigma_{\mathrm{SR}}(G)$, is defined as

$$\sigma_{\mathrm{SR}}(G) := \mathrm{sp}(G) \Big/ \mathbb{E}\left[\mathrm{sp}\left(\hat{G}\right)\right],$$

if $\mathbb{E}\left[\mathrm{sp}\left(\hat{G}\right)\right] > 0$, and $\sigma_{\mathrm{SR}}(G) := 1$ otherwise.

Remark 1. If $\mathbb{E}\left[\mathrm{sp}\left(\hat{G}\right)\right] = 0$ then it is necessarily the case that $\mathrm{sp}(G) = 0$, since it is always true that $\mathbb{P}\left(\hat{G} = G\right) > 0$. Moreover, if $\mathrm{sp}(G) = 0$ and $\mathbb{E}\left[\mathrm{sp}\left(\hat{G}\right)\right] = 0$ then the number of simplicial pairs is as expected and so we define $\sigma_{\mathrm{SR}}(G) = 1$.

Remark 2. We have mentioned already that the sizes of the edges in a simplicial pair are important. For this reason, we condition on $\hat{G} \sim \mathrm{CL}(G)$ having no multiset edges (edges containing multiple instances of the same vertex).

Remark 3. We approximate $\mathbb{E}\left[\mathrm{sp}\left(\hat{G}\right)\right]$ rather than compute this expectation exactly. For a graph G, computing $\mathbb{E}\left[\mathrm{sp}\left(\hat{G}\right)\right]$ is quite difficult as we discuss in the open problems presented in the journal version of this paper.

Recalling Example 1, we have that $\mathrm{sp}(G_1) = 6$ and $\mathbb{E}\left[\mathrm{sp}\left(\hat{G}\right)\right] \approx 4.3$, meaning $\sigma_{\mathrm{SR}}(G_1) \approx 1.4$, whereas $\mathrm{sp}(G_2) = 0$ and $\mathbb{E}\left[\mathrm{sp}\left(\hat{G}_2\right)\right] \approx 0.2 > 0$, meaning $\sigma_{\mathrm{SR}}(G_2) = 0$. Thus, the simplicial ratio can clearly distinguish G_1 and G_2.

2.2 The simplicial matrix

For a graph G, write $\mathrm{sp}(G, i, j)$ for the number of simplicial pairs (e_1, e_2) in G with $|e_1| = i$ and $|e_2| = j$ with $i < j$. Then, letting $\hat{G} \sim \mathrm{CL}(G)$ conditioned on having no multiset edges, the simplicial matrix of G, denoted by $M_{\mathrm{SR}}(G)$, is the partial matrix with cell (i, j) equalling

$$M_{\mathrm{SR}}(G, i, j) := \frac{\mathrm{sp}(G, i, j)}{\mathbb{E}\left[\mathrm{sp}\left(\hat{G}, i, j\right)\right]}$$

whenever $i < j$ and G contains edges of size i and of size j (and substituting 0 if there are no simplicial pairs of this type), and with cell (i, j) being empty otherwise.

We will see in Section 3 that the simplicial matrix reveals information about real-world graphs that the simplicial ratio alone does not. In particular, a hypothesis we make in this paper, as suggest by these matrices, is that *the composition of an edge in a real-world network becomes more dependent on simpliciality as the edge size increases.*

Let us again revisit Examples 1. We have

$$M_{\mathrm{SR}}(G_1) \approx \begin{bmatrix} \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \mathbf{3.8} & \mathbf{1.7} & \mathbf{1} \\ \emptyset & \emptyset & \emptyset & \emptyset & \mathbf{2.4} & \mathbf{1} \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \mathbf{1} \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \end{bmatrix} \qquad M_{\mathrm{SR}}(G_2) \approx \begin{bmatrix} \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \\ \emptyset & \emptyset & \emptyset & \mathbf{0} & \mathbf{0} \\ \emptyset & \emptyset & \emptyset & \emptyset & \mathbf{0} \\ \emptyset & \emptyset & \emptyset & \emptyset & \emptyset \end{bmatrix}.$$

The simplicial matrix for G_1 unpacks the information about its simplicial interactions. Indeed, the simplicial ratio simply tells us that the number of simplicial pairs is 1.4 times more than expected. On the other hand, the simplicial matrix tells us that all 3 simplicial pairs involving the edge of size 6 are to be expected, whereas the other three simplicial pairs are at least somewhat surprising.

3 Empirical results

We compute the simplicial ratio and simplicial matrix for the same 10 graphs that were analysed in [17]. The graphs are taken from [16] and full descriptions can be found there. The datasets are

- three proximity-based networks: contact-primary-school, contact-high-school and, hospital-lyon,
- two email networks: email-enron and email-eu,
- three biological networks: diseaseome, disgenenet and ndc-substances, and
- two misc. networks: congress-bills and tags-ask-ubuntu.

For each graph, we restrict to edges of sizes 2, 3, 4 and 5, and we restrict to the largest connected component if the graph is not connected. Additionally, we throw away multi-edges.

3.1 The data

We first show Table 1 which includes the simplicial ratios, as well as the number of vertices and edges, for each graph. Then, in Figure 2 we show the simplicial matrix of each graph. For readability we show only the non-empty cells of the partial matrices.

G	$\lvert V(G)\rvert$	$\lvert E(G)\rvert$	$[\lvert E_2\rvert, \lvert E_3\rvert, \lvert E_4\rvert, \lvert E_5\rvert]$	$\sigma_{\mathrm{SR}}(G)$
disgenenet	469	232	$[53, 77, 61, 41]$	15.99
ndc-substances	1468	2661	$[973, 695, 505, 488]$	10.30
diseasome	372	256	$[131, 80, 23, 22]$	6.46
contact-h.s.	326	2680	$[1733, 842, 99, 6]$	6.16
email-enron	142	1325	$[808, 316, 138, 63]$	5.20
congress-bills	1236	1455	$[526, 365, 316, 248]$	4.88
email-eu	958	21307	[13k, 5k, 2k, 1k]	4.74
contact-p.s.	242	2480	$[997, 1364, 116, 3]$	1.89
hospital-lyon	75	1535	$[947, 539, 48, 1]$	0.97
tags-ask-ubuntu	3021	145053	[28k, 52k, 39k, 25k]	0.69

Table 1. The simplicial ratio of 10 real networks and the corresponding bottom-up simplicial ratio and top-down simplicial ratio for the 7 temporal networks. The graphs are ordered according to $\sigma_{\mathrm{SR}}(G)$, from largest to smallest.

3.2 Analysis

Simplicial ratio We see that that biology networks are, on average, more surprisingly simplicial than contact-based networks and email networks. In contrast, it was shown in [17] that contact-based networks are the closest to their simplicial closures and biological networks are furthest from theirs. In fact, comparing the ranks of the 3 existing measures (sf, es, fes) and the ranks from our simplicial ratio (sr), we get the following Kendall correlation values.

	sf	es	fes	sr
sf	1.000	0.706	0.989	-0.539
es	0.706	1.000	0.722	-0.535
fes	0.989	0.722	1.000	-0.556
sr	-0.539	-0.535	-0.556	1.000

These values show that our ranking system is, quite substantially, negatively correlated with the ranking systems in [17]. While there are likely many factors contributing to this negative correlation, one strong factor is *edge density*. Indeed, the three biology networks, disgenenet, ndc-substances, and diseasome, are very sparse graphs, with disgenenet and diseasome having fewer edges than vertices. On the other hand, the three contact based graphs, contact-h.s., contact-p.s., and hospital-lyon, are quite dense.

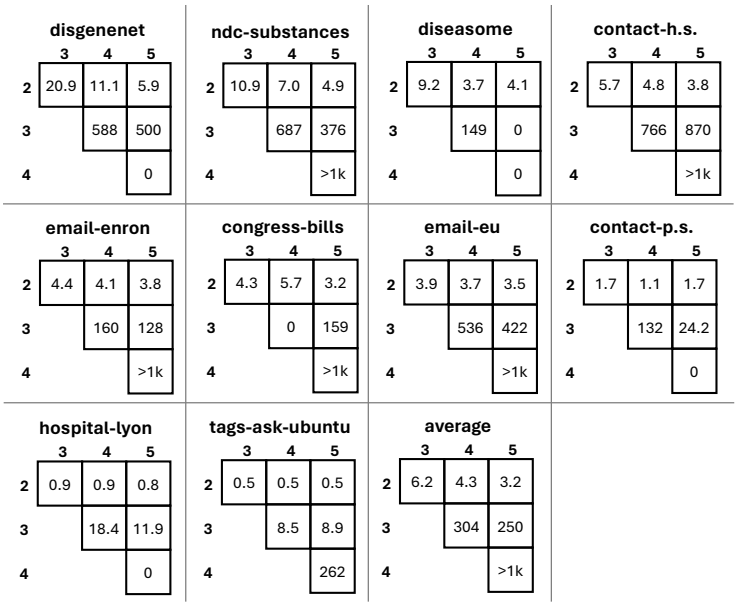

Fig. 2. The simplicial matrix of 10 real networks, as well as the cell-wise average matrix, restricted to edges of size 2, 3, 4, and 5. For each graph G, only non-empty cells of $M_{\mathrm{SR}}(G)$ are shown. The value of a cell is replaced with "> 1k" whenever the value is above 1000.

Simplicial matrix An immediate take-away from these matrices is that simplicial interactions become more surprising as edge size increases. Although this feature is interesting, there is at least a partial explanation for this phenomenon, namely, that sparse Chung-Lu graphs (and many other sparse random graphs with independent edge generation) are quite unlikely to generate any simplicial pairs unless at least one of the edges involved is of size 2.

4 A model that incorporates simpliciality

We define a random graph model, called the *simplicial Chung-Lu model*, that generalizes the Chung-Lu hypergraph model defined in [10]. We present the model in full detail in the journal version of this paper. Here, we give a high level summary of the model.

1. Along with the usual input parameters of the Chung-Lu model, we also require a parameter $q \in [0, 1]$.
2. Iteratively to construct an edge, we decide if it is a "normal" edge, or a simplicial edge, based on a q-weighted coin flip.
   - If the coin flip is successful, we construct a new edge e by sampling a previously constructed edge e' and creating a simplicial pair (e, e').
   - If the coin flip is unsuccessful, we construct a new edge as per the usual Chung-Lu model.

5 Experiments

5.1 Descriptions of the experiments

We perform two experiments on the 10 real networks and on the corresponding simplicial Chung-Lu graphs for varying $q \in \{0, 0.5, 1\}$. The experiments are outlined below.

Giant component growth with random edge selection: We choose a uniform random order for $E(G)$ and track the size of the largest component as edges are added to G according to a random ordering.

Information diffusion from a single source: We initialize a function $w_0 : V(G) \to [0, 1]$ with $w_0(v) = 0$ for all vertices, except for one randomly chosen vertex v^* which has $w_0(v^*) = 1$. Then, in round $i + 1$, we choose a random edge e and, for each $v \in e$, set $w_{i+1}(v) = w(e)/|e|$, where $w(e) = \sum_{u \in e} w(u)$ (keeping $w_{i+1}(v) = w_i(v)$ for all $v \notin e$). We track the Wasserstein-1 distance (also known as the "earth mover's distance" [22]) between w_i and the uniform distribution $w_\infty : V(G) \to 1/|V(G)|$.

Insisting on connected graphs These experiments, and in particular the two diffusion experiments, are highly dependent on connectivity. The real graphs are restricted to their largest component, and so we insist that the random graphs are also connected. To achieve this, we modify the simplicial Chung-Lu model and insist that incoming edges must connect disjoint components, until the point the graph is connected when we continue generating edges as normal. A full description of this algorithm is presented in the journal version of this paper.

5.2 The results

Here, we will show the results for the two graphs: **hospital-lyon** and **disgenenet**. Recall that the **hospital-lyon** graph has a simplicial ratio of approximately 0.97, whereas the **disgenenet** graph has a ratio of approximately 15.99. The full collection of results can be found in the journal version.

Experiment 1: random growth In this first experiment we see the following. For **hospital-lyon** the real graph grows in a near identical way to the random model with $q = 0$ and $q = 0.5$, whereas the random model with $q = 1$ grows much slower. In contrast, for **disgenenet** the real graph grows most similarly to the random model with $q = 1$ whereas the random model with $q = 0.5$ grows slightly faster, and for $q = 0$ even faster still. Of course, these graphs have very different growth behaviour due to the difference in edge densities. Nevertheless, this result suggests that the high simplicial ratio of **disgenenet** plays a role in slowing down the growth of the graph, whereas the low simplicial ratio of **hospital-lyon** leads it to grow as quickly as a classical Chung-Lu model.

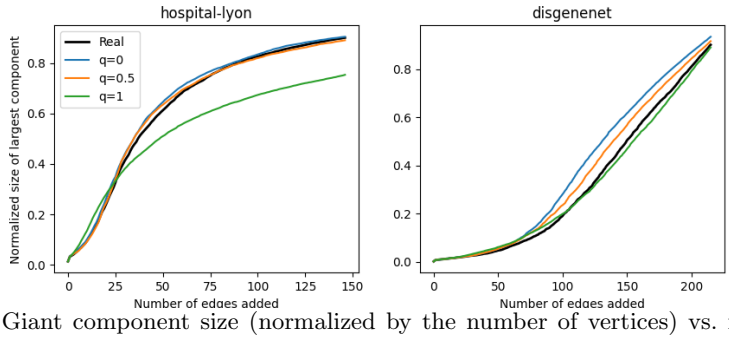

Fig. 3. Giant component size (normalized by the number of vertices) vs. number of edges added in the random growth process on the **hospital-lyon** graph (left) and the **disgenenet** graph (right). The curve is the point-wise average across 10000 independent experiments: for the real graph the edges are resampled each time, and for the random models the entire graphs are resampled each time.

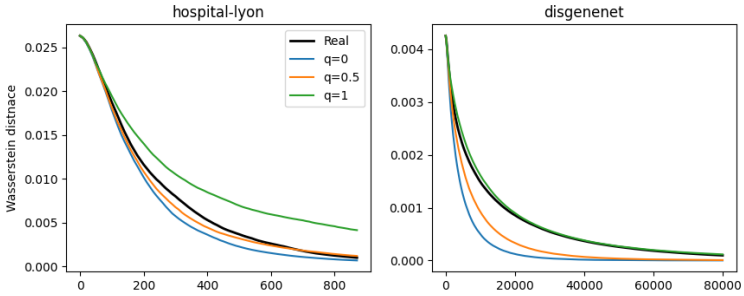

Fig. 4. Wasserstein distance to uniform vs. number of rounds in the single-source diffusion process on the **hospital-lyon** graph (left) and the **disgenenet** graph (right). The curve is the point-wise average across 10000 independent experiments: for the real graph the chosen edges per round, as well as the location of the initial vertex with weight 1, are resampled each time, and for the random models the entire graphs are resampled each time.

Experiment 2: single-source diffusion This experiment suggests that information diffusion is slower on highly simplicial graphs vs. non-simplicial graphs. We note, however, that the diffusion process on **hospital-lyon** is still slower than that of a random model with $q = 0.5$. Surely there are more features of this real graph not captured by random models that contribute to the slower diffusion time.

Acknowledgements

This research is supported in part by the Tutte Institute for Mathematics and Computing (TIMC), a division of the Communications Security Establishment (CSE). TIMC does not specifically endorse the contents of this work, or any other work by the authors. Any opinions or positions represented herein do not represent the official position of CSE or the Government of Canada.

References

1. Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.G., Petri, G.: Networks beyond pairwise interactions: structure and dynamics. Physics Reports **874**, 1–92 (2020)
2. Benson, A.R., Abebe, R., Schaub, M.T., Jadbabaie, A., Kleinberg, J.: Simplicial closure and higher-order link prediction. Proceedings of the National Academy of Sciences **115**(48), E11,221–E11,230 (2018)
3. Benson, A.R., Gleich, D.F., Higham, D.J.: Higher-order network analysis takes off, fueled by classical ideas and new data. arXiv preprint arXiv:2103.05031 (2021)
4. Benson, A.R., Gleich, D.F., Leskovec, J.: Higher-order organization of complex networks. Science **353**(6295), 163–166 (2016)
5. Easley, D., Kleinberg, J.: Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge university press (2010)
6. Feng, S., Heath, E., Jefferson, B., Joslyn, C., Kvinge, H., Mitchell, H.D., Praggastis, B., Eisfeld, A.J., Sims, A.C., Thackray, L.B., et al.: Hypergraph models of biological networks to identify genes critical to pathogenic viral response. BMC bioinformatics **22**(1), 287 (2021)
7. Jackson, M.O.: Social and economic networks. Princeton university press (2010)
8. Juul, J.L., Benson, A.R., Kleinberg, J.: Hypergraph patterns and collaboration structure. arXiv preprint arXiv:2210.02163 (2022)
9. Kamiński, B., Misiorek, P., Prałat, P., Théberge, F.: Modularity based community detection in hypergraphs. In: International Workshop on Algorithms and Models for the Web-Graph, pp. 52–67. Springer (2023)
10. Kamiński, B., Poulin, V., Prałat, P., Szufel, P., Théberge, F.: Clustering via hypergraph modularity. PloS one **14**(11), e0224,307 (2019)
11. Kaminski, B., Prałat, P., Théberge, F.: Mining complex networks. Chapman and Hall/CRC (2021)
12. Kamiński, B., Prałat, P., Théberge, F.: Hypergraph artificial benchmark for community detection (h–abcd). Journal of Complex Networks **11**(4), cnad028 (2023)
13. Kamiński, B., Misiorek, P., Prałat, P., Théberge, F.: Modularity based community detection in hypergraphs (2024). URL https://arxiv.org/abs/2406.17556
14. Kim, J., Lee, D.S., Goh, K.I.: Contagion dynamics on hypergraphs with nested hyperedges. Physical Review E **108**(3) (2023). DOI 10.1103/physreve.108.034313. URL http://dx.doi.org/10.1103/PhysRevE.108.034313
15. Lambiotte, R., Rosvall, M., Scholtes, I.: Understanding complex systems: From networks to optimal higher-order models. arXiv preprint arXiv:1806.05977 (2018)
16. Landry, N.W., Lucas, M., Iacopini, I., Petri, G., Schwarze, A., Patania, A., Torres, L.: XGI: A Python package for higher-order interaction networks. Journal of Open Source Software **8**(85), 5162 (2023). DOI 10.21105/joss.05162. URL https://joss.theoj.org/papers/10.21105/joss.05162
17. Landry, N.W., Young, J.G., Eikmeier, N.: The simpliciality of higher-order networks. EPJ Data Science **13**(1), 17 (2024)
18. LaRock, T., Lambiotte, R.: Encapsulation structure and dynamics in hypergraphs. Journal of Physics: Complexity **4**(4), 045,007 (2023). DOI 10.1088/2632-072X/ad0b39. URL https://dx.doi.org/10.1088/2632-072X/ad0b39
19. Lee, G., Bu, F., Eliassi-Rad, T., Shin, K.: A survey on hypergraph mining: Patterns, tools, and generators (2024). URL https://arxiv.org/abs/2401.08878
20. Newman, M.: Networks. Oxford university press (2018)

21. Odda, T.: On properties of a well-known graph or what is your ramsey number. Annals of the New York Academy of Sciences **328**, 166 – 172 (2006). DOI 10.1111/j.1749-6632.1979.tb17777.x
22. Rubner, Y., Tomasi, C., Guibas, L.: A metric for distributions with applications to image databases. In: Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271), pp. 59–66 (1998). DOI 10.1109/ICCV.1998.710701
23. Tian, H., Zafarani, R.: Higher-order networks representation and learning: A survey. arXiv preprint arXiv:2402.19414 (2024)
24. Torres, L., Blevins, A.S., Bassett, D., Eliassi-Rad, T.: The why, how, and when of representations for complex systems. SIAM Review **63**(3), 435–485 (2021). DOI 10.1137/20M1355896
25. Zhang, Y., Lucas, M., Battiston, F.: Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. Nature Communications **14**(1) (2023). DOI 10.1038/s41467-023-37190-9