Supplementary Information for: Explaining Twitter's inability to reduce vaccine misinformation during the COVID-19 pandemic

Supplementary Information

We constructed a simulation, based on [1], to better articulate how a system's architecture can facilitate or inhibit a social media platform's control efforts. Social media platforms are very large, hosting hundreds of millions or even billions of daily active users. Directly measuring and modeling networks of this size is prohibitive. Furthermore, the specific network structure of any given social media platform is constantly changing. The aim of this simulation is to rank different platform architectures regarding the extent to which they might facilitate or inhibit control of information flow.

Since we cannot directly measure *orders of growth* of different social media platform architectures. Our aim in doing so is to be able to assess whether a platform whose design follows a given architecture is more or less flexible — and therefore controllable — than another platform following a different architecture. Our simulation is based on Moses' Theory of Generic System Architectures [1], which quantifies the *flexibility* of different structures by counting up the total number of paths in the corresponding network. Flexibility is important because Moses theorizes that it is inversely related to its controllability [2]. The rationale is that if individual paths or nodes are disrupted, e.g., because content or accounts are removed, other accounts on the same platform can still access interdicted information through the use of alternate pathways. When the nodes represent social media accounts that are operating in a manner that is not aligned with the platform owner (e.g., when anti-vaccine accounts post vaccine misinformation despite a platform's policy prohibiting this misinformation), this flexibility can be used to undermine the platform owner's control efforts because attempts to remove posts or accounts can be circumvented by relying on the same alternate paths that facilitate flexibility.

Ideally, we would simulate several networks with the aim of calculating each one's total number of alternate pathways through which information might flow. However, exhaustive enumeration of all paths in a graph is known to be a #P-complete problem [3, 4], meaning that this enumeration can only be calculated precisely by brute force. This is prohibitively computationally expensive for large graphs. We therefore approximate the total number of paths in a graph using its communicability [5] – a widely-used proxy that is more computationally tractable [6, 7, 8, 9]. Specifically, the total network communicability $C^{TN}(A)$ for each graph can be calculated by computing the exponential of its adjacency matrix,

$$[e^{(A)}]_{i,j} = \sum_{k=1}^{\infty} \frac{[A^k]_{i,j}}{k!} \tag{1}$$

and then summing across all nodes.

$$C^{TN}(A) = \mathbf{1}^T(e^A)\mathbf{1} \tag{2}$$

Our approach is to estimate the *order of growth* of the total network communicability of networks associated with different social media platform architectures, and then to compare these across architectures. To do so, we simulated networks associated with several of the generic system architectures identified by Moses [1] varying the total number of nodes in each network. In practice, total network communicability follows similar orders of growth as those derived for number of paths in prior work [1].

Social Media Platform Architectures

We claim that Twitter's system architecture undermines its controllability. Beyond empirical evidence for this claim, the simulation in this section compares networks resembling Twitter's structure to networks with structures that result from other architectures.

Layered Hierarchies

A platform's architecture may place constraints on how information is allowed to flow between accounts. One such constraint, used by Facebook [10], is known as “layering”. In layered structures, nodes are divided into a set of layers arranged in a hierarchy such that nodes in layer n may only connect to nodes in layers $n-1$, n, or $n+1$. Furthermore, the internal structure of a layered hierarchy need not be specified [1]. Moses [1, 2, 11] theorizes that layering can make inflexible systems more flexible, but can also make overly-flexible (and hence uncontrollable) systems more controllable. For each architecture described below, we also examined how segmenting the resulting networks into two or three layers affected their communicability. Although this need not be the case in general, for the purposes of this simulation, we assume that each layer has the same structure. For example, a three-layered scale-free structure with 2400 nodes is made up of three scale-free networks with 800 nodes each, with connections between nodes in adjacent layers. We examined the effects of segmenting each of the structures described below into 2- and 3-layered hierarchies, holding the total number of nodes constant.

Generic System Architectures

For purposes of comparison, we generated several network structures based on the generic system architectures defined in [1].

Regular Trees

One of the most restrictive system architectures is a regular tree structure, which requires that each node has exactly one parent and all but the leaf nodes have a fixed number of children. Although social media platforms are never this restrictive in practice, platforms such as Reddit or online discussion boards have some tree-like properties such as a requirement that messages can only be seen by followers of a given subreddit and, in some cases, are subject to removal by a moderator. Thus, tree structures represent a limiting case.

We generated tree-structured graphs using the IGRAPH python package [12] examining values of the tree branching factors (number of children per parent node) ranging from 2 through 9. For each branching factor, we generated trees with depth ranging from 2 through 9 and retained all trees with at most 2400 nodes. We calculated the total network communicability for each network and found that, holding branching factor fixed, communicability grows linearly with the total number of nodes, similar to the relationship shown in prior work [1] (see Table 1).

Regular Lattices

A somewhat less restrictive system architecture is a regular lattice structure, in which each node has exactly d neighbors. This structure is also quite unrealistic for social media platforms. However, what it has in common with Twitter is the absence of an explicit hierarchy. We therefore include it for comparison purposes.

We generated lattices using the IGRAPH python package [12] with the dimensionality (total number of neighbors) of each node ranging from 2 through 5 retaining all lattices with at most 2400 nodes. We calculated the total network communicability for each network and found that, holding lattice dimensionality fixed, communicability grows in an inverse power relationship with the total number of nodes per lattice dimension, as shown in prior work [1] (see Table 2). Since the number of nodes in the network is equal to the number of nodes per dimension raised to the power of d, the overall order of growth of these lattices is linear, albeit higher than trees.

Regular Toruses

A variant of the lattice structure is the regular torus, in which all edges of the lattice "wrap around". This structure, although still unrealistic for real social media platforms, is nevertheless the starting point for the Watts-Strogatz model, which is intended to induce the small world effect after random rewiring [13]. We therefore include this model because it is an example of an architecture that induces community structure.

We generated toruses using the IGRAPH python package [12] with the dimensionality (total number of neighbors) of each node ranging from 2 through 5 retaining all lattices with at most 2400 nodes. We calculated the total network communicability for each network and found that, holding torus dimensionality fixed, communicability grows linearly with the total number of nodes in the torus (see Table 4) and, in practice, has very similar communicability values to those found in regular lattices for single-layered structures; however, as the number of layers increases toruses appear to be

Table 1: Communicability Values for Tree-Structured Hierarchies, evaluated at n=3,686,697 – the total number of accounts in our data after applying 3-core network decomposition to identify communities.

Regular Trees; $O(n)$		
One Layer		
Branching Factor	Equation	$log_{10}C^{TN}$
2	$C^{TN} = 9.88n - 23.39$	7.56
3	$C^{TN} = 13.02n - 51.73$	7.68
4	$C^{TN} = 16.92n - 95.45$	7.80
5	$C^{TN} = 21.63n - 152.93$	7.90
6	$C^{TN} = 27.63n - 285.37$	8.01
7	$C^{TN} = 33.45n - 295.78$	8.09
8	$C^{TN} = 41.85n - 474.30$	8.19
9	$C^{TN} = 51.97n - 742.20$	8.28
Two Layers		
2	$C^{TN} = 26.85n - 127.17$	8.00
3	$C^{TN} = 35.37n - 266.52$	8.12
4	$C^{TN} = 46.00n - 518.94$	8.23
5	$C^{TN}C = 58.79n - 831.40$	8.34
6	$C^{TN} = 72.07n - 975.90$	8.42
7	$C^{TN} = 90.93n - 1608.03$	8.53
8	$C^{TN} = 113.76n - 2578.58$	8.62
9	$C^{TN} = 141.27n - 4035.02$	8.72
Three Layers		
2	$C^{TN} = 39.53n - 280.89$	8.16
3	$C^{TN} = 52.07n - 588.66$	8.28
4	$C^{TN} = 67.08n - 954.74$	8.39
5	$C^{TN} = 86.57n - 1836.33$	8.50
6	$C^{TN} = 106.12n - 2155.46$	8.59
7	$C^{TN} = 133.89n - 3551.66$	8.69
8	$C^{TN} = 167.51n - 5695.32$	8.79
9	$C^{TN} = 208.82n - 8912.15$	8.88

Table 2: Communicability Values for Regular Lattice Structures, evaluated at n=3,686,697 – the total number of accounts in our data after applying 3-core network decomposition to identify communities.

	Lattices; $O(n)$	
Dimensionality	Equation	$log_{10} C^{TN}$
	One Layer	
2	$C^{TN} = (7.39\sqrt[d]{n} - 10.26)^d$	8.30
3	$C^{TN} = (7.34\sqrt[d]{n} - 9.91)^d$	9.15
4	$C^{TN} = (7.16\sqrt[d]{n} - 9.21)^d$	9.94
5	$C^{TN} = (6.89\sqrt[d]{n} - 8.46)^d$	10.63
	Two Layers	
2	$C^{TN} = (17.35\sqrt[d]{n} - 2.76)^d$	9.04
3	$C^{TN} = (14.09\sqrt[d]{n} - 8.40)^d$	10.01
4	$C^{TN} = (15.56\sqrt[d]{n} - 18.54)^d$	11.29
5	$C^{TN} = (16.92\sqrt[d]{n} - 24.03)^d$	12.55
	Three Layers	
2	$C^{TN} = (25.86\sqrt[d]{n} - 5.06)^d$	9.39
3	$C^{TN} = (18.70\sqrt[d]{n} - 12.48)^d$	10.38
4	$C^{TN} = (18.97\sqrt[d]{n} - 22.61)^d$	11.63
5	$C^{TN} = (27.50\sqrt[d]{n} - 46.07)^d$	13.58

Table 3: Communicability Values for Regular Torus Structures, evaluated at n=3,686,697 – the total number of accounts in our data after applying 3-core network decomposition to identify communities.

	Toruses; $O(n)$	
Dimensionality	Equation	$log_{10}C^{TN}$
	One Layer	
2	$C^{TN} = 54.60n - 9.39$	8.30
3	$C^{TN} = 403.78n - 502.27$	9.17
4	$C^{TN} = 2999.49n - 17994.72$	10.04
5	$C^{TN} = 22540.47n - 455929.19$	10.92
	Two Layers	
2	$C^{TN} = 148.46n - 73.81$	8.74
3	$C^{TN} = 1099.47n - 3761.86$	9.61
4	$C^{TN} = 8233.00n - 127165.56$	10.48
5	$C^{TN} = 61271.35n - 2478687.96$	11.35
	Three Layers	
2	$C^{TN} = 218.61n - 163.03$	8.91
3	$C^{TN} = 1618.94n - 8308.83$	9.78
4	$C^{TN} = 12122.82n - 280870.77$	10.65
5	$C^{TN} = 90220.08n - 5474681.77$	11.52

Table 4: Communicability Values for Regular Torus Structures, evaluated at n=3,686,697 – the total number of accounts in our data after applying 3-core network decomposition to identify communities.

Dimensionality	Toruses; $O(n)$ Equation	$log_{10}C^{TN}$
	One Layer	
2	$C^{TN} = 54.60n - 9.39$	8.30
3	$C^{TN} = 403.78n - 502.27$	9.17
4	$C^{TN} = 2999.49n - 17994.72$	10.04
5	$C^{TN} = 22540.47n - 455929.19$	10.92
	Two Layers	
2	$C^{TN} = 148.46n - 73.81$	8.74
3	$C^{TN} = 1099.47n - 3761.86$	9.61
4	$C^{TN} = 8233.00n - 127165.56$	10.48
5	$C^{TN} = 61271.35n - 2478687.96$	11.35
	Three Layers	
2	$C^{TN} = 218.61n - 163.03$	8.91
3	$C^{TN} = 1618.94n - 8308.83$	9.78
4	$C^{TN} = 12122.82n - 280870.77$	10.65
5	$C^{TN} = 90220.08n - 5474681.77$	11.52

less flexible, and hence more controllable, than lattices.

Complete Graphs

Trees, lattices, and toruses are all relatively sparse structures. In contrast, many social media clusters are strongly connected. An extreme version of this observation yields an architecture with a fully-connected "team" or "complete" structure, in which each node is connected to all of its neighbors.

We generated fully-connected graphs using the IGRAPH python package [12] with the total number of nodes ranging from 100 through 700 nodes in increments of 100 (larger numbers of nodes had communicability values that were so high that they led to an overflow error). We calculated the total network communicability for each network and found that communicability grows exponentially with the total number of nodes in the network (see Table 5).

Table 5: Communicability Values for Complete Structures, evaluated at n=3,686,697 – the total number of accounts in our data after applying 3-core network decomposition to identify communities.

	Teams; $O(e^n)$	
Number of Layers	Equation	$log_{10}C^{TN}$
1	$C^{TN} = 13.94e^n$	3,705,048.33
2	$C^{TN} = 163.67e^{\frac{n}{2}}$	1,849,776.64
3	$C^{TN} = 361.51e^{\frac{n}{3}}$	1,233,186.90

Scale-Free Networks

Twitter does not impose any restrictions on which accounts may follow other accounts. Thus, it possesses an unrestricted architecture, which Moses [11] simply refers to as a "network". Absent any restrictions, a significant body of empirical work has demonstrated that such networks tend to self-organize into highly-clustered communities in a manner that approximates a scale-free network [14, 15, 16]. We therefore calculated the communicability of several scale free networks and compared the communicability of these structures to those resulting from "generic architectures" defined in prior work [1].

To generate scale free networks, we used the IGRAPH python package [12] to create graphs with degree distributions matching our data. Specifically, we fit a power law to the empirical degree distribution of our dataset using the POWERLAW python package [17], and estimated the power law exponent of this dataset as $\gamma = 1.46$. We next randomly generated a degree sequence following this power law degree distribution, and used this degree sequence to randomly generate a scale-free network. Given this network, we next introduced community structure into our network by generating 25 simulated communities, each of which were created using the Watts-Strogatz model [13] with an average degree of 1 and a rewiring probability of 0.10. Next, we randomly selected 50 pairs of nodes and added edges between them if they were not already adjacent and if they both had at least two edges.

We examined networks with 150, 300, 600, 1200, and 2400 nodes, simulating 1000 networks for each network size. For each simulated network, we calculated its total network communicability and extracted the median and 95% confidence bounds (i.e., the 2.5 and 97.5 percentile communicability values). We compared the communicability of these scale free networks to those associated with Moses' generic architectures (see Table 6).

Table 6: Communicability Values for Scale-Free Networks with Average Degree = 1 and γ = 1.46 evaluated at n=3,686,697 – the total number of accounts in our data after applying 3-core network decomposition to identify communities.

Scale-Free Networks; $O(e^{\frac{n}{\gamma}})$		
Percentile	Equation	$log_{10}C^{TN}$
	One Layer	
2.5	$C^{TN} = 0.92e^{\frac{0.70n}{\gamma}}$	24,583.67
50.0	$C^{TN} = 62.94e^{\frac{0.77n}{\gamma}}$	26,289.57
97.5	$C^{TN} = 6518.29e^{\frac{0.85n}{\gamma}}$	28,008.56
	Two Layers	
2.5	$C^{TN} = 17.70e^{\frac{0.37n}{\gamma}}$	15,756.60
50.0	$C^{TN} = 139.47e^{\frac{0.41n}{\gamma}}$	16,889.05
97.5	$C^{TN} = 2020.98e^{\frac{0.45n}{\gamma}}$	18,183.69
	Three Layers	
2.5	$C^{TN} = 37.94e^{\frac{0.25n}{\gamma}}$	12,241.88
50.0	$C^{TN} = 209.23e^{\frac{0.28n}{\gamma}}$	13,035.62
97.5	$C^{TN} = 1143.98e^{\frac{0.31n}{\gamma}}$	14,075.13

Simulation results show that the total network communicability of scale-free networks grows as order $O(e^{\frac{n}{\gamma}})$, which is larger than all but the highest order of growth of all generic architectures. In contrast, platforms such as Facebook, which utilize layering, are likely more controllable given the same number of nodes. Although we leave a systematic exploration of multiple different platforms to future work, we observe that Reddit, which utilizes elements of tree-like structures, may among the most controllable of social media platforms.

References

[1] Broniatowski, D. A. & Moses, J. Measuring flexibility, descriptive complexity, and rework potential in generic system architectures. *Systems Engineering* **19**, 207–221 (2016).

[2] Moses, J. *Flexibility and Its Relation to Complexity and Architecture*, 197–206 (Springer Berlin Heidelberg, 2010).

[3] Valiant, L. G. The complexity of enumeration and reliability problems. *siam Journal on Computing* **8**, 410–421 (1979).

[4] Roberts, B. & Kroese, D. Estimating the number of st paths in a graph. *Journal of Graph Algorithms and Applications* **11**, 195–214 (2007).

[5] Estrada, E. & Hatano, N. Communicability in complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* **77**, 036111 (2008).

[6] Benzi, M. & Klymko, C. Total communicability as a centrality measure. *Journal of Complex Networks* **1**, 124–149 (2013).

[7] Estrada, E. & Hatano, N. Communicability graph and community structures in complex networks. *Applied Mathematics and Computation* **214**, 500–511 (2009).

[8] Estrada, E. The communicability distance in graphs. *Linear Algebra and its Applications* **436**, 4317–4328 (2012).

[9] De la Cruz Cabrera, O., Jin, J., Noschese, S. & Reichel, L. Communication in complex networks. *Applied Numerical Mathematics* **172**, 186–205 (2022).

[10] Broniatowski, D. A., Simons, J. R., Gu, J., Jamison, A. M. & Abroms, L. C. The efficacy of facebook's vaccine misinformation policies and architecture during the covid-19 pandemic. *Science Advances* **9**, eadh2132 (2023).

[11] Moses, J. *The Anatomy of Large Scale Systems* (2012).

[12] Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **Complex Systems**, 1695 (2006).

[13] Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *nature* **393**, 440–442 (1998).

[14] Barabási, A.-L. Scale-free networks: A decade and beyond. *Science* **325**, 412–413, DOI: 10.1126/science.1173299 (2009).

[15] Barabási, A.-L. & Bonabeau, E. Scale-free networks. *Scientific american* **288**, 50–9 (2003).

[16] Broido, A. D. & Clauset, A. Scale-free networks are rare. *Nature communications* **10**, 1017 (2019).

[17] Alstott, J., Bullmore, E. & Plenz, D. powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one* **9**, e85777 (2014).