# Gregory J. Clark

Research Statement

**My research is in pure and applied discrete mathematics; specifically, spectral hypergraph theory, computational social science, and graph statistics.** In my work I develop hypergraph models and apply them to empirical phenomena. As part of my research I collaborate with corporate partners (e.g., Google, Meta, Twitter, et al.) to solve future-facing industry problems in a measurable, explainable, and verifiable way. These projects have lead to actionable insights for thought leaders, peer-reviewed research papers, and policy shaping reports. In this statement I will describe my work in discrete mathematics and then discuss how it relates to my industry focused research in network analysis.

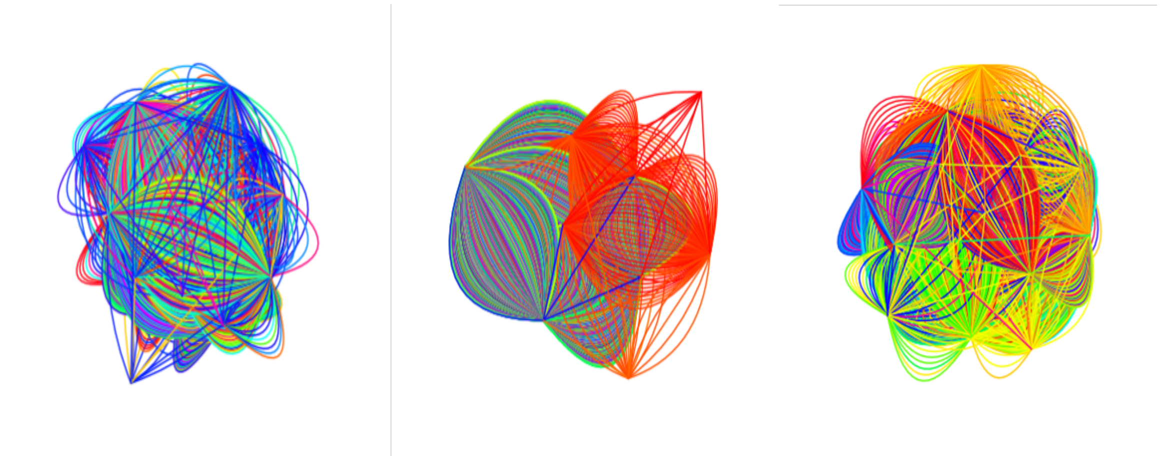

Figure 1: Hypergraph models of cross platform communication about Covid-19 topics.

A central theme of my research is exploring the characteristic polynomial of a hypergraph. A *hypergraph* is a generalization of the traditional graph (i.e., network) where edges are allowed to connect more than two vertices. Examples of hypergraphs used to model social networks addressing Covid-19 topics are given in Figure 1 (higher dimensional edges are monochromatic). This work is connected to numerous fields of mathematics because the aforementioned polynomial is the resultant of a system of multilinear homogeneous equations. The resultant, which is a generalization of the determinant, is central to algebraic geometry, commutative algebra, and numerical multilinear algebra among other areas. In particular, my dissertation gives a generalization of the Harary-Sachs Theorem by providing a combinatorial description of the coefficients of characteristic polynomial of a hypergraph [1]. This result is valuable because we can compute the leading coefficients of this particular polynomial without needing to compute all the coefficients, which is NP-hard to compute in general. Furthermore, I have provided a numerically stable algorithm which can compute the characteristic polynomial of a hypergraph given its set spectrum and leading coefficients. This allows for the computation of the characteristic polynomial of a hypergraph when traditional tools from commutative algebra (i.e., the resultant) have been insufficient. I have further applied this work to industry sourced problems by examining how the principal eigenvector of a hypergraph and its shadows can vary.

# 1 Spectral Hypergraph Theory

Given a graph $G$ we would like to understand its structure. One can define a certain matrix of a graph (e.g., adjacency, incidence, Laplacian, etc.) and connect the multiset of eigenvalues of this matrix back to the graph. For example, the existence of strongly regular graphs (i.e., a regular graph where all pairs of adjacent vertices have $a$ common neighbors and all pairs of non-adjacent vertices have $b$ common neighbors) is sharply constrained by the graph's spectrum. Famously, the Hoffman-Singleton theorem, whose proof integrally uses a delicate analysis of a strongly regular graph's spectral properties, says that the only graphs with girth 5 and diameter 2 are necessarily $d$-regular for $d \in \{2, 3, 7, 57\}$. A construction for the $d = 57$ case remains open. Spectral graph theory has also found applications to real-world problems, for example, in the identification of key users in a social network via the eigencentrality measure. This idea is central to Google's PageRank patent. With the recent surge of interest in data science, network analysts are considering more complex sets of data. In many cases one may want to capture salient properties of a network where connections can be better expressed through group connections. To do so, we extend these results to hypergraphs.

My research focuses on spectral hypergraph theory, where we relate the structure of a hypergraph to its spectrum and vice versa. A hypergraph is a generalization of a graph wherein edges potentially contain more than two vertices. Hypergraphs allow one to model complex systems with greater fidelity: committees of representatives, similarities between friends, and the evolution of a network over time. The cost of analyzing this richer structure is paid for in theoretical and computational complexity. Where the adjacency characteristic polynomial of a graph can be quickly computed as the determinant of a matrix, the adjacency characteristic polynomial of a hypergraph is the resultant of a multilinear homogeneous system of equations [2] (which is known to be NP-hard to compute in general [3]). As such, my primary line of research seeks to understand the adjacency characteristic polynomial of a hypergraph. To this end we have generalized the Harary-Sachs Theorem to hypergraphs [1], characterized the spectrum of hypertrees [4], and provided a numerically stable algorithm for computing the characteristic polynomial of a hypergraph given its set of eigenvalues and some leading coefficients [5]. Below I discuss the motivation, impact, and continuing research agenda in spectral hypergraph theory.

## 1.1 The Generalized Harary-Sachs Theorem for Hypergraphs

An early, seminal result in spectral graph theory of Harary [6] (and later, more explicitly, Sachs [7]) expressed the coefficients of a graph's characteristic polynomial as a certain weighted sum of the counts of various subgraphs of $G$.

**Theorem 1.** *(Harary-Sachs Theorem) Let $G$ be a labeled simple graph on $n$ vertices. If $H_i$ denotes the collection of $i$-vertex graphs whose components are edges or cycles, and $c_i$ denotes the codegree-$i$ coefficient of the characteristic polynomial of $G$ (i.e., the coefficient of $\lambda^{n-i}$), then*

$$c_i = \sum_{H \in H_i} (-1)^{c(H)} 2^{z(H)} [\#H \subseteq G]$$

*where $c(H)$ is the number of components of $H$, $z(H)$ is the number of components which are cycles, and $[\#H \subseteq G]$ denotes the number of (labeled) subgraphs of $G$ which are isomorphic to $H$.*

The Harary-Sachs Theorem relates the spectrum of a graph and its elementary subgraphs (i.e., disjoint union of edges and cycles). We have generalized this theorem to hypergraphs and provided an analogous description of elementary subgraphs which we refer to as *Veblen graphs*. This result allows us to compute partial information about the characteristic polynomial of a hypergraph when computing the whole polynomial is computationally costly. We are excited to have generalized the Harary-Sachs Theorem to hypergraphs in the following way.

**Theorem 2.** *([1]) Let $\mathcal{H}$ be a $k$-uniform hypergraph on $n$ vertices. If $\mathcal{V}_i$ denotes the set of $k$-uniform Veblen multi-hypergraphs (i.e., all vertices have degree divisible by $k$), and $c_i$ denotes the codegree-$i$ coefficient in the characteristic polynomial of $\mathcal{H}$, then*

$$c_i = \sum_{H \in \mathcal{V}_i(\mathcal{H})} (-(k-1)^n)^{c(H)} C_H(\#H \subseteq \mathcal{H})$$

*where $c(H)$ is the number of components of $H$, $C_H$ is a certain computable coefficient of $H$, and $(\#H \subseteq \mathcal{H})$ is the number of particular maps of $H$ to subgraphs of $\mathcal{H}$.*

Theorem 2 is a faithful generalization of the Harary-Sachs Theorem as it simplifies to the Harary-Sachs Theorem when $\mathcal{H}$ is a graph (i.e., a 2-uniform hypergraph) [8]. We point out that this result implies that the spectrum of a hypergraph is computable from the counts of its Veblen subgraphs, just as the counts of its elementary subgraphs determine the spectrum of a graph. While this situation for hypergraphs is predictably more complicated than the graph case, it does directly connect the characteristic polynomial of a hypergraph to its structure.

## 1.2 The Spectrum of Hypertrees

When studying the spectrum of a graph, a simple family to consider is the collection of trees. As an example, one can apply the Harary-Sachs Theorem to show that the coefficients of the characteristic polynomial of a tree count the number of matchings of a particular size. This yields that multiplicity of the zero eigenvalue of a tree is equal to the size of its largest matching. Surprisingly, a similar bound on the multiplicity of the zero eigenvalue for an arbitrary graph continues to elude description. We discuss our results concerning the spectrum of a hypertree which are stronger than their tree analogue. We demonstrate the peculiarity of this dichotomy by providing a spectral characterization of so-called "power trees".

In [4] we show that the spectrum of a hypertree is the collection of all totally non-zero eigenvalues of its subtrees. An eigenvalue is *totally non-zero* if the eigenvalue is non-zero and corresponds to an eigenvector with all non-zero entries. In [9] the authors show that the totally non-zero eigenvalues of a hypertree are roots of a certain matching polynomial. **Using their formula we characterize the set spectrum of a hypertree as the union of the totally non-zero eigenvalues of its induced subtrees**. In other words, we show that the totally non-zero eigenvalues of a hypertree are necessarily eigenvalues of any hypertree which contains it as a subgraph. Note that this is a variant of the Cauchy Interlacing Theorem which says that the eigenvalues of a subgraph (formed by removing one vertex) interlace the eigenvalues of the original graph. This is somewhat surprising in light of the fact that the same statement is not true for ordinary graphs. We demonstrate this peculiarity by considering power trees. A *power tree* is a $k$-uniform hypergraph created by adding $k-2$ new vertices to each edge of a tree. We have shown that a power tree is characterized by its eigenvalues being cyclotomic.

## 1.3 Stably Computing the Multiplicity of Known Roots

The characteristic polynomial of a $k$-uniform hypergraph with $n$ vertices is the univariate (in $\lambda$) polynomial obtained from the resultant of a family of $n$ multilinear homogeneous polynomials of degree $k-1$, minus $\lambda$ times a diagonal form of the same degree. By properties of the resultant, the degree of the characteristic polynomial is $n(k-1)^{n-1}$. Computing the resultant is known to be NP-hard over any field, in general [3]. Thus, computing the characteristic polynomial of a hypergraph using traditional tools from commutative algebra is intractable. However, we can try to determine the characteristic polynomial of a hypergraph another way. Given the set of roots of a polynomial without multiplicity and an appropriate number of leading coefficients one can determine the multiplicity of its roots using the Faddeev-LeVerrier algorithm, a matrix form of the Newton Identities. **In [5] we provide a numerically stable algorithm for computing the multiplicity of the roots of a polynomial where the roots (without multiplicity) and some leading coefficients are known.** The algorithm is stable in the sense that if an eigenvalue is approximated by an $\varepsilon$-disk, where $\varepsilon$ depends "reasonably" on the parameters of the problem, the resulting disk approximating its multiplicity contains exactly one integer. Our bound on $\varepsilon$ is "reasonable" in that the number of bits required to approximate each root is proportional to the number of distinct roots of $p$ and the logarithms of the ratio of the smallest difference of the roots with the largest difference of roots, the largest root, and the largest coefficient. The crux of the algorithm is a method to invert a Vandermonde matrix via a special factorization, when direct inversion would be numerically unstable.

We apply this algorithm to compute the adjacency characteristic polynomial of various hypergraphs. By Theorem 2 we can determine the leading coefficients of the characteristic polynomial, so it remains to determine the set of eigenvalues of the hypergraph. We can apply the aforementioned results to determine the spectrum (without multiplicities) of a hypertree. For a general hypergraph we can appeal to a method of Lu and Man to determine a subset of the set spectrum [10]. As an example, we compute the characteristic polynomial of the *Rowling hypergraph*, as shown in Figure 2, to be

$$\begin{aligned}\phi(\mathcal{R}) =& x^{133}(x^3-1)^{27}(x^{15}-13x^{12}+65x^9-147x^6+157x^3-64)^{12}\\ &\cdot (x^6-x^3+2)^6(x^6-17x^3+64)^3.\end{aligned}$$

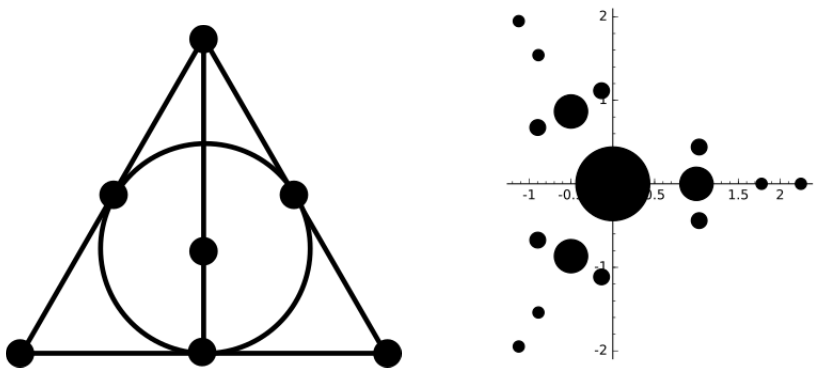


Figure 2: The Rowling hypergraph and its spectrum where a disk is centered at each root and its area is proportional to the root's multiplicity.

## 1.4 Future Work

**Quasirandom Hypergraphs:** I will apply Theorem 2 to prove a sufficient spectral condition for quasirandom hypergraphs. Intuitively, a hypergraph is *quasirandom* if it has the same number of copies of a particular subgraph as one would expect in a random graph (where, in its simplest form, each edge is taken with probability 1/2). This idea was first introduced for graphs in [11] and was later extended to hypergraphs in [12]. Generally speaking, the idea is that if one can show that a graph satisfies a particular condition then it is quasirandom. In [12] the authors show that a hypergraph is quasirandom if it has approximately the expected number of even partial octahedra (as described therein). Using Theorem 2 we aim to restate this condition in terms of the coefficients and perhaps the spectrum itself by showing that the linear combinations of subgraph counts appearing in the result are indeed "forcing sets" for quasirandomness.

**Multiplicity of the Zero Eigenvalue:** One can apply the Harary-Sachs Theorem to show that the multiplicity of the zero eigenvalue of a tree is equal to the size of its largest matching. I plan to prove a similar statement for hypertrees using the fact that the eigenvalues of a hypertree are the roots of a certain matching polynomial. I believe it is also possible to use Theorem 2 to show, by collecting summands corresponding to the same subgraph, that a hypergraph has a *coefficient threshold* which provides an upper-bound on the multiplicity of the zero eigenvalue. By determining the coefficient threshold of hypertrees and other classes of hypergraphs we could provide an upper-bound, or perhaps an explicit formula, for the multiplicity of the zero eigenvalue.

**Open Source Software:** Computing the adjacency characteristic polynomial of a hypergraph is NP-hard, in general. We have provided a numerically stable algorithm for computing the characteristic polynomial of a hypergraph given its set of eigenvalues and an tractable number of leading coefficients. I will continue working on each facet of this endeavor. This includes finding faster ways to compute the leading coefficients and store this information in an open access database, expanding on the Lu-Man method to provide an algorithm for determining eigenvalues of a larger class of hypergraphs, and utilizing high performance computing resources to perform these computations. In particular, I aim to compute the characteristic polynomial of the Fano Plane, especially since I know its first fifteen coefficients! I further plan on making this algorithm open source as a service to the mathematical community.

# 2 Hypergraph Models of Empirical Phenomena

The analysis of relational data and their corresponding networks has become an integral part to the study of complex systems [13, 14, 15, 16]. Graph theory (and more generally, network analysis) is commonly employed in these situations because of its suite of statistical techniques. A prevailing example is network centrality. A typical application of network centrality involves locating influential users in a network. For example, quantifying the effects of collaboration networks, the advantages of network position, and the resilience of one's reputation. Centrality measures are also a common tool in scientific fields such as machine learning, biology, and engineering.

There are several centrality measures to consider but the most frequently used are degree, betweeness, closeness, and eigenvector centrality. While the aforementioned notions of centrality are related, they can be different in practice. Notably, the Krakhardt kite (see Figure 3) is an example of a network where different vertices are identified as being the most central under degree (vertex-3), betweeness (vertex-7), and closeness centrality (vertex-5 and vertex-6). The Krackhardt kite demonstrates that different definitions of centrality can lead to the identification of different nodes as being the most central. **As multi-dimensional data becomes increasingly ubiquitous in data science, it is natural to wonder if a centrality measure (like eigenvector centrality) can similarly identify different nodes as being the most central depending on the dimension of the data. Remarkably, the answer is in the affirmative.** We present an example of this phenomenon and discuss its implications below.

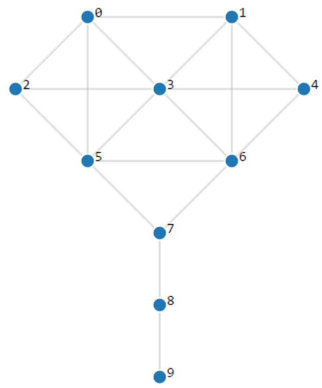


Figure 3: The Krackhardt kite.

## 2.1 The Effect of Network Dimensionality on Eigenvector Centrality

Historically, multi-dimensional data has been projected down to lower dimensions to simplify analysis. In particular, data which could be represented as a hypergraph is projected down to a traditional, two-dimensional graph. This operation, referred to as the *shadow* (aka clique expansion in some literature), replaces each edge with a clique of the same size. For example, the edge $\{1,2,3\}$ would become $\{\{1,2\},\{1,3\},\{2,3\}\}$ under the shadow operation. In Figure 4 we present an example of a *bow tie* hypergraph, where each edge is incident to three nodes, and its corresponding shadow graph. It is natural to ask to what extent can the principal eigenvector of a hypergraph and its shadow vary? One approach is to consider their *spectral ranking*, that is, the ordering of the vertices from most central to least central.

We have shown that the bow tie hypergraph $\mathcal{B}_1$, given in Figure 5, and its shadow disagree between who is more central: vertex-1 or vertex 4? In the 3-dimensional 'hypergraph case', vertex-1 is more central than vertex-4. However, in the 2-dimensional 'shadow' case, the opposite is true. In practice, this means that the 'most influential users of a social network' can vary depending on the dimension of the model. **What's more, we can extend this example to a hypergraph where the most central node, under the eigenvector centrality measure, changes under the shadow operation (see Figure 5).** As a result, networks which are created from projecting higher dimensional data down to 2-dimensions can lead to different identifications of the most central node, even when no information is lost.

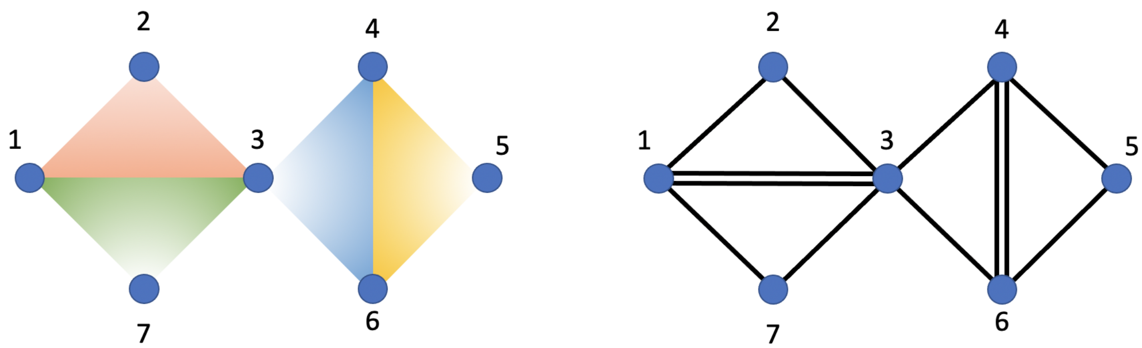


Figure 4: A 3-uniform bow tie hypergraph $\mathcal{B}_1$ (left) and its shadow $\partial\mathcal{B}_1$ (right).

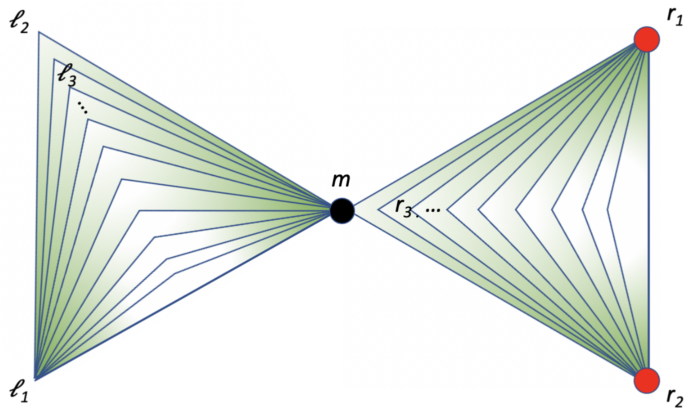


Figure 5: The 8-pleated bow tie, $\mathcal{B}_8$. The vertices $\{r_1, r_2\}$ are the most central vertices (under the principal eigenvector centrality) but $m$ is the most central in the shadow.

## 2.2 The Correlation of Shadow Eigenvectors

In the previous section we discussed how the spectral ranking of a hypergraph and its shadow can vary. One may also consider the extent to which these rankings may differ globally (as opposed to local inversions). To answer this question we consider how the eigenvectors of a hypergraph and its shadow can correlate. It is commonly accepted that the principal eigenvector and degree vector of a graph have a high Pearson correlation. In this case, we are interested in their Spearman correlation (i.e., the Pearson correlation of the rank function). Amazingly, the Spearman correlation of the principal eigenvector of $\mathcal{B}_8$, given in Figure 5, and its shadow is 0.07110. Moreover, this is *exact*. A summary of the correlation between various vectors associated to $\mathcal{B}_1$ and $\mathcal{B}_8$ are presented in Table 1.

Table 1: The Pearson $r_p$ and Spearman $r_s$ correlation coefficients of $y$, the principal eigenvector of a hypergraph $H$; $x$, the principal eigenvector of its shadow $\partial^2(H)$; and $d$, the normalized degree vector of $H$.

| Hypergraph | $r_p(y,x)$ | $r_p(y,d)$ | $r_p(x,d)$ | $r_s(y,x)$ | $r_s(y,d)$ | $r_s(x,d)$ |
|---|---|---|---|---|---|---|
| $\mathcal{B}_8$ | 0.99455 | 0.99601 | 0.99985 | **0.07110** | 0.72623 | 0.73439 |
| $\mathcal{B}_1$ | 0.99549 | 0.98784 | 0.99230 | 0.88889 | 0.94281 | 0.94281 |

## 2.3 Classifying Expert Opinion from Text Data

**The *Here to be Heard* campaign is the result of a global survey of over 10,000 woman from 88 countries with the aim of helping women achieve their full potential.** The study was lead by Mars, Incorporated with partnership from the Oxford Future of Marketing Initiative. I discuss the details of the collaboration and my role in it in my Diversity Statement. Participants were asked one question: *what needs to change so that more women can reach their full potential?*. The results from this report are being used to reshape Mars to empower women within the company and their global community. This survey speaks to the larger problem of representation within decision making groups at various levels of organisations around the world. In particular, can those without a particular lived experience adequately make decisions with only outside information? We hypothesize that this is not the case.

To test our hypothesis we re-frame the question as a classification problem. We define an expert to be someone with a related lived experience. For example, an expert on the aforementioned question is anyone who has identified as a woman at some point in their life. The question then is whether we can classify an expert by the topics used in their response alone.

To address this problem we utilized natural language processing and latent Dirichlet allocation topic modeling to identify the key topics for experts globally. Through this process we identified 28 primary topics which were mentioned by experts in their responses. We then worked with qualitative analysts to group these topics into eight major themes. With these topics in hand, we explored how they were presented across different demographics.

A surprising result from this analysis was the difference between the frequency of topic mentions between experts and nonexperts. The RMSE of responses varied by 0.0643, in other words, expert and non-expert respondents mentioned the same topics with roughly the same frequency on a global scale. This is evidence for a behavioural phenomenon known as *parroting*. Parroting is a process by which non-experts repeat expert opinion without the depth of understanding [17, 18]. To better gauge the depth of knowledge, we shifted our attention to how respondents formed narratives from these topics. To this end I constructed a hypergraph association model to explore the complexity of responses by demographic as shown in Figure 6. Once again we were surprised that both expert and non-expert respondents contained roughly the same number of topics in their responses on average. However, an interesting network structure emerged: the density of expert and non-expert association networks were drastically different. To be precise, expert respondents formed a nearly complete network (all pairs of associations were accounted for) while non-expert respondents only accounted for 70% of all possible associations. In summary, non-experts were able to mimic the responses of experts but were not able to form the same associations between topics.

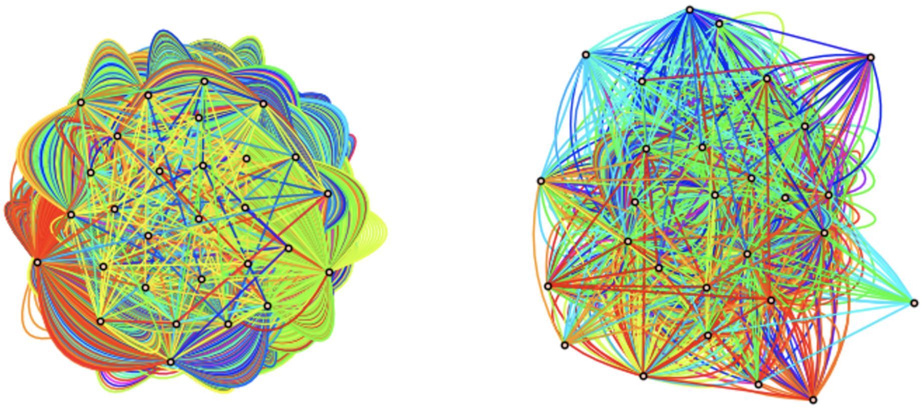

Figure 6: Hypergraph model of expert (left) and non-expert (right) respondents. Each monochromatic clique represents the topics used in a given response.

## 2.4 Future Work

**Testing for spectral rank inversions:** As demonstrated above, the spectral ranking of a hypergraph and its shadow can differ. We say that such a hypergraph is *umbral*. There are various constructions of umbral hypergraphs; however, there is no test for whether a given hypergraph is umbral or not. An initial family to consider is that of $k$-cylinders (i.e., $k$-colorable $k$-uniform hypergraphs [2]) because of their prevalence in modeling heterogeneous data and structural simplicity. I will provide a characterisation of hypergraphs which are umbral and further create a test for determining if a $k$-cylinder is umbral which is feasible (i.e., computationally $O(n^k)$).

**Co-opting social movements:** In the digital age, grassroot social movements emerge to redefine cultural norms. This process involves the co-opting of various groups and ideologies to gain traction in the mainstream media. One vehicle for this is hashtags (#). Hashtags create awareness and discussion, spread ideas, better affiliate individuals with a community, integrate resources from the Internet, facilitate policy formation, social change, and democracy [19, 20]. As a social movement attracts more attention it welcomes further co-opting which can unintentionally shift its focus. We will study social phenomena related to this concept by examining emergent topics in the life cycle of a social movement.

**Quantifying corporate authenticity:** We have considered the classification of text data as either expert or non-expert opinion. I am interested in studying how generalizable these results are to other contexts. In particular, if we can classify client facing corporate communications arising reflect expert or non-expert opinion. For example, can we determine if a firm's purpose statement is authentic (c.f., green washing)? I will create an open source software package for researchers to apply this methodology to enable social science researchers.

# 3 Irregularity Measures and Their Applications

There are several measures of graph irregularity which are used to determine how 'close' a given graph is to being regular. One such measure, the *principal ratio*, is the ratio of the greatest and least entry of the principal eigenvector of a graph $G$. We reserve $(\lambda, x)$ to denote the principal eigenpair of the adjacency matrix of $G$. The principal ratio, then, is denoted $\gamma(G) = x_{\max}/x_{\min}$. A limitation of this statistic is its sensitivity to the position (and to an extent the degree) of individual vertices in the graph. For example, its applicability to empirical networks where the principal eigenvector is highly dispersed (e.g., scale-free networks [21, 22, 23]). This motivates the study of the underlying vector's dispersion.

Given a vector $x \in \mathbb{R}^n_+$ let $\mu$ denote its mean and $\sigma$ its standard deviation. The coefficient of variation of $x$ is the ratio $\sigma/\mu$. For convenience we will consider the square of the coefficient of variation which we denote

$$c_x := \left(\frac{\sigma}{\mu}\right)^2.$$

The coefficient of variation and its square appear frequently across multiple disciplines: economic inequality [24], reliability of measures [25], and efficiency of experimental designs [26]. Interestingly, the inverse of the coefficient of variation, $\mu/\sigma$, is the signal-to-noise ratio (SNR). The SNR is used informally to measure the ratio of useful to irrelevant information in a conversation [27]. It has further found use in information theory [28], optics [29], internet topology [30], medical imaging [31], and nuclear engineering [32].

## 3.1 Distribution of Graphical Vectors

Studying the distribution of vectors associated to graphs is not new. Notably the variance of the degree vector was an early irregularity measure [33]. As such, we consider the coefficient of variation of the degree vector in addition to the principal eigenvector. We reserve $d$ to denote the degree vector of a graph $G$. We denote the square of the coefficient of variation of the principal eigenvector as $c_e$ and the degree vector as $c_d$. Studying the distribution of these two vectors in tandem will further aid in motivating their use in practice.

We show that $c_d$ and $c_e$ are both bounded above by the same function of principal ratio (namely, $\gamma^2 - 1$). We further consider various graph families and compute the limit of the aforementioned statistics of said graphs. **We present graph families for which the limit**

**of their coefficient of variation is 0, non-zero, or diverges to infinity. We summarize our results in Table 2 and note that all six cases for the limit of the principal ratio and the coefficient of variation (for the principal eigenvector or degree vector, respectively) are achieved.**

Table 2: The limit of $\gamma^2 - 1$, $c_e$, and $c_d$ for various graph families including the complete graph with an edge removed; a particular complete triparite graph, a complete split graph (i.e., $S(n,m) = K_{n+m} - K_m$); the complete graph with a pendant edge, a kite whose head is an $r$-regular graph; the star; and the Cartesian product of a graph with itself.

| $G_n$ | $\lim \gamma^2 - 1$ | $\lim c_e$ | $\lim c_d$ |
|---|---|---|---|
| $K_n - K_2$ | 0 | 0 | 0 |
| $K_{1,n,n}$ | 3 | 0 | 0 |
| $S(n,kn)$ | $\left(\frac{\sqrt{4k+1}+1}{2}\right)^2 - 1$ | $\frac{k\left(\frac{\sqrt{4k+1}-1}{k}-2\right)^2}{(\sqrt{4k+1}+1)^2}$ | $\frac{k^3}{(2k+1)^2}$ |
| $P_2K_{n-1}$ | $\infty$ | 0 | 0 |
| $P_nG_n^r$ | $\infty$ | $-$ | $\left(\frac{r-2}{r+2}\right)^2$ |
| $K_{1,n}$ | $\infty$ | 1 | $\infty$ |
| $G^{\square n}$ | $\infty$ | $\infty$ | 0 |

## 3.2 Divergence of Local and Global Clustering Coefficient

Consider the complete split graph $S(n,m) = K_{n,m} \cup K_n$. We have shown that the principal ratio, eigenvector dispersion, and degree dispersion of $S(n,nk)$ all converge to a positive number for fixed $k$. In the case when $k=1$, we showed that the principal ratio of $S(n,n)$ converges to $\varphi$, the golden ratio. We further show that this family has the rare property that the average clustering coefficient and transitivity diverge for $S(n,kn)$ as we take $n$ and $k$ to be arbitrarily large. To the best of our knowledge this is the third such graph family to have this property [34] (c.f., the friendship (aka windmill, $n$-fan) graph [21] and the agave graph [35]). Notably, this property was observed in brokerage networks formed by illicit marketplaces which are closely approximated by complete split graphs [36, 37, 38].

The *Watts-Strogatz clustering coefficient* (see [13]) of a vertex $i$ is a measure of the transitivity of local connections in a network

$$C_i = \frac{2t_i}{d_i(d_i-1)}$$

where $t_i$ is the number of triangles which contain $i$. The *average Watts-Strogatz clustering coefficient* of a graph $G$ is defined

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i.$$

Finally the *transitivity* of a graph (see [39, 40]) is defined to be

$$T = \frac{3(\#\text{ of triangles})}{\sum_{i=1}^{n}\binom{d_i}{2}}.$$

Using this language we show

**Theorem 3.**

$$\lim_{k\to\infty}\lim_{n\to\infty} \overline{C}(S(n,kn)) = 1 \textit{ and } \lim_{k\to\infty}\lim_{n\to\infty} T(S(n,kn)) = 0.$$

### 3.3 Future Work

**Extremal Graph:** The kite graph (aka lollipop graph) is the extremal graph for the maximum hitting time of a graph as well as the maximum principal ratio for graphs on $n$ vertices [41, 42, 43]. In each case, the head size (i.e., $s$ in $P_mK_s$) is a function of $n$. The maximum hitting time occurs when $s \approx (2n+1)/3$ while maximising the principal ratio requires $s \approx n/\log n$. We conjecture that a kite graph maximises $c_e$ when the head size is precisely $s = 4$. This conjecture was previously observed in the context in molecular graphs [44]. I will prove this conjecture in the positive which will resolve open problems in various fields.

**Eigenvector with Small Variance but Large Degree Variance:**. In Table 2 we provide the limit of the principal ratio, the coefficient of variation of the principal eigenvector, and the coefficient of variation of the degree vector. Our initial results compared the limit of the principal ratio with that of the coefficients of variation. Note an example of a graph family $G_n$ where $\lim c_e(G_n) = 0$ but $\lim c_d(G_n) = \infty$ remains open. I conjecture and will prove that no such family exists.

**Coefficient of Variation in Practice:** As previously discussed the coefficient of variation arises in numerous fields and contexts. We have constructed various graph families for which the limit $c_e$ and $c_d$ can converge or diverge. I will review the literature concerning applications of the coefficient of variation and examine if the conclusions apply in the case of the principal eigenvector, degree vector, or both.

# References


[1] Gregory J. Clark and Joshua N. Cooper. A Harary-Sachs theorem for hypergraphs. *Journal of Combinatorial Theory, Series B*, 149:1–15, 2021.

[2] Joshua Cooper and Aaron Dutle. Spectra of uniform hypergraphs. *Linear Algebra Appl.*, 436(9):3268–3292, 2012.

[3] Bruno Grenet, Pascal Koiran, and Natacha Portier. The multivariate resultant is NP-hard in any characteristic. In *Mathematical foundations of computer science 2010*, volume 6281 of *Lecture Notes in Comput. Sci.*, pages 477–488. Springer, Berlin, 2010.

[4] Gregory J. Clark and Joshua Cooper. On the adjacency spectra of hypertrees. *Electronic Journal of Combinatorics*, 25, 2017.

[5] Gregory J. Clark and Joshua N. Cooper. Stably computing the multiplicity of known roots given leading coefficients. *Numerical Linear Algebra with Applications*, 27(2):e2275, 2020.

[6] Frank Harary. The determinant of the adjacency matrix of a graph. *SIAM Rev.*, 4:202–210, 1962.

[7] Horst Sachs. Über Teiler, Faktoren und charakteristiche Polynome von Graphen. *Teil I*, 12:7–12, 1966.

[8] Gregory J. Clark and Joshua Cooper. Applications of the harary-sachs theorem for hypergraphs, 2021.

[9] Wei Zhang, Liying Kang, Erfang Shan, and Yanqin Bai. The spectra of uniform hypertrees. *Linear Algebra Appl.*, 533:84–94, 2017.

[10] Linyuan Lu and Shoudong Man. Connected hypergraphs with small spectral radius. *Linear Algebra Appl.*, 509:206–227, 2016.

[11] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 12 1989.

[12] Fan Chung. Quasi-random hypergraphs revisited. *Random Structures & Algorithms*, 40(1):39–48, 2011.

[13] Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

[14] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[15] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

[16] Ronald S Burt. *Structural holes: The social structure of competition*. Harvard university press, 2009.

[17] Palash Kamruzzaman. Civil society or comprador class, participation or parroting? *Progress in Development Studies*, 13(1):31–49, 2013.

[18] Joost Van Spanje and Nan Dirk De Graaf. How established parties reduce other parties electoral support: the strategy of parroting the pariah. *West european politics*, 41(1):1–27, 2018.

[19] Hong Tien Vu, Lei Guo, and Maxwell E McCombs. Exploring the world outside and the pictures in our heads a network agenda-setting study. *Journalism & mass communication quarterly*, 91(4):669–686, 2014.

[20] Bridget Tombleson and Katharina Wolf. Rethinking the circuit of culture: How participatory culture has transformed cross-cultural communication. *Public Relations Review*, 43(1):14–25, 2017.

[21] Béla Bollobás and Oliver M Riordan. Mathematical results on scale-free random graphs. *Handbook of graphs and networks: from the genome to the internet*, pages 1–34, 2003.

[22] Fan Chung, Linyuan Lu, and Van Vu. Eigenvalues of random power law graphs. *Annals of Combinatorics*, 7(1):21–33, 2003.

[23] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.

[24] Anthony B Atkinson et al. On the measurement of inequality. *Journal of economic theory*, 2(3):244–263, 1970.

[25] Orit Shechtman. The coefficient of variation as an index of measurement reliability. In *Methods of clinical epidemiology*, pages 39–49. Springer, 2013.

[26] Frank Ephraim Grubbs. *Statistical measures of accuracy for riflemen and missile engineers*. FE Grubbs, 1991.

[27] Will Gragido, Johnl Pirc, Nick Selby, and Daniel Molina. Chapter 4 - signal-to-noise ratio. In Will Gragido, Johnl Pirc, Nick Selby, and Daniel Molina, editors, *Blackhatonomics*, pages 45 – 55. Syngress, Boston, 2013.

[28] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.

[29] Albert Rose. *Vision: human and electronic*. Springer Science & Business Media, 2013.

[30] Ravi Kumar, P Ragbavan, Sridhar Rajagopalan, and Andrew Tomkins. The web and social networks. *Computer*, 35(11):32–36, 2002.

[31] WA Edelstein, GH Glover, CJ Hardy, and RW Redington. The intrinsic signal-to-noise ratio in nmr imaging. *Magnetic resonance in medicine*, 3(4):604–618, 1986.

[32] David I Hoult and RE Richards. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *Journal of Magnetic Resonance (1969)*, 24(1):71–85, 1976.

[33] Francis K Bell. A note on the irregularity of graphs. *Linear Algebra and its Applications*, 161:45–54, 1992.

[34] Ernesto Estrada. When local and global clustering of networks diverge. *Linear Algebra and its Applications*, 488:249–263, 2016.

[35] Ernesto Estrada and Eusebio Vargas-Estrada. Distance-sum heterogeneity in graphs and complex networks. *Applied Mathematics and Computation*, 218(21):10393–10405, 2012.

[36] Paolo Spagnoletti, Federica Ceci, and Bendik Bygstad. Online black-markets: An investigation of a digital infrastructure in the dark. *Information Systems Frontiers*, pages 1–16, 2021.

[37] Felipe Thomaz. The digital and physical footprint of dark net markets. *Journal of International Marketing*, 28(1):66–80, 2020.

[38] Felipe Thomaz and John Hulland. Abnormal structure of digital black markets. *In preparation.*, 2022.

[39] Mark EJ Newman. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2):404–409, 2001.

[40] Stanley Wasserman, Katherine Faust, et al. *Social network analysis: Methods and applications*. Cambridge university press, 1994.

[41] Graham Brightwell and Peter Winkler. Maximum hitting time for random walks on graphs. *Random Structures & Algorithms*, 1(3):263–276, 1990.

[42] Michael Tait and Josh Tobin. Characterizing graphs of maximum principal ratio. *The Electronic Journal of Linear Algebra*, 34:61–70, 2018.

[43] Lele Liu and Changxiang He. The maximum principal ratio of graphs. *Linear and Multilinear Algebra*, pages 1–16, 2022.

[44] Gerta Rücker, Christoph Rücker, and Ivan Gutman. On kites, comets, and stars. sums of eigenvector coefficients in (molecular) graphs. *Zeitschrift für Naturforschung A*, 57(3-4):143–153, 2002.