 doi:10.20944/preprints202103.0715.v1

A Set-theoretic Approach to Modeling Network Structure

John L. Pfaltz
Dept. of Computer Science, University of Virginia
jlp@virginia.edu

Abstract

Three computer algorithms are presented. One reduces a network $\mathcal{N}$ to its interior, $\mathcal{I}$. Another counts all the triangles in the network, and the last randomly generates networks similar to $\mathcal{N}$ given just its interior $\mathcal{I}$. But these algorithms are not the usual numeric programs that manipulate a matrix representation of the network; they are set-based. `Union` and `meet` are essential binary operators; `contained_in` is the basic relational comparator.

The interior $\mathcal{I}$ is shown to have desirable formal properties and to provide an effective way of revealing "communities" in social networks.

1 Introduction

The text book way of describing network structure is to represent a network, $\mathcal{N}$, as two sets (N, L) where N is a set of **nodes**, and L is a set of unordered pairs $\{x, y\} \subseteq N$, called **links**[1] [1, 6]. But, although text book network theory is almost always set based, virtually all computer network algorithms are algebraic [14, 15]. This is most understandable. Any network can be represented by its adjacency matrix, $A_{n,n}$, where $a_{i,j} = 1$ if $\{i, j\}$ is a link and 0 otherwise. There is an abundance of matrix algorithms one can use, in contrast to the dearth of practical set manipulation software.

To overcome this problem, we created our own C^{++} set management system [16]. In it, sets are strongly typed; for example there are "sets of nodes" and "sets of links" which are completely distinct. Invoking the subroutines that execute set operations can be awkward and takes time to master; but one can faithfully duplicate all of the pseudocode presented in this paper.[2]

In Section 2, we develop the notion of a network's "interior" and present a scalable algorithm to compute it. This subset $\mathcal{I} \subseteq \mathcal{N}$ captures almost all of the essential features of a network $\mathcal{N}$ as is illustrated in Section 4. Moreover, $\mathcal{I}$ is an interesting mathematical structure in its own right. We demonstrate that it is a system of chordless cycles which is a well-defined algebraic matroid. But, we do not further develop this aspect.

[1] In graph theory these unordered pairs are called "edges". This seems to be derived from the edges of the solid "dodecahedron puzzle" of Sir William Hamilton (1857) and retained through inertia. However, since in social networks they connect individuals, it seems more appropriate to call them "links".

[2] C^{++} source code for all procedures of this paper can be obtained from the author.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2021 doi:10.20944/preprints202103.0715.v1

In Section 3, we explore a number of the network properties that have been discussed in the literature. In particular, we present a rather simple algorithm to count the triangles in a network, and show how $\mathcal{I}$ can reveal the "community" structure of a social network.

Section 4 is somewhat unusual. To support the claim that the interior, $\mathcal{I}$, is a good descriptor of $\mathcal{N}$, we present a procedure `expand` which randomly generates networks "like" $\mathcal{N}$, based solely on $\mathcal{I}$. Presumably, one network is "like" another if the properties discussed in Section 3 are similar.

2 The Interior

Let S be a set. An **operator** $\tau : 2^S \rightarrow 2^S$ is an injective function which maps subsets of S into subsets of S. We denote operators by greek letters and use postfix notation, as in $Y.\tau$, where $Y \subseteq S$. An operator φ said to be a **closure** operator if for all $X, Y \subseteq S$, (C1) $Y \subseteq Y.\varphi$ (expansive), (C2) $X \subseteq Y$ implies $X.\varphi \subseteq Y.\varphi$ (monotone), and (C3) $Y.\varphi.\varphi = Y.\varphi$ (idempotent). Closure operators are a staple of topological mathematics.

If we replace axiom C1 with an contractive axiom I1, so that for all $X, Y \in \mathcal{S}$, (I1) $Y.\iota \subseteq Y$ (contractive), (I2) $X \subseteq Y$ implies $X.\iota \subseteq Y.\iota$ (monotone), (I3) $Y.\iota.\iota = Y.\iota$ (idempotent), then ι is said to be an **interior** operator. We use ι to denote interior operators and φ to denote closure operators; they are similar, except that one is contractive while the other is expansive.

If one visualizes S as a polytope, then its closure might be the smallest sphere containing $\mathcal{S}$ (often called its convex hull), while its interior could be the largest inscribed sphere, or ball. Alternatively, if one thinks of S as being a bit of irregular surface terrain with ridges and valleys, then a closure operator fills in the valleys until the terrain is uniformly smooth. An interior operator, in contrast, levels the peaks and ridges until a smooth terrain is obtained.

Let $\mathcal{N}$ be a network. For any $Y \subseteq N$, we say the **neighborhood** of Y is $Y.\eta = \{z | \exists y \in Y, \{y, z\} \in L\} \cup Y$. (In graph theory, $Y.\eta$ is sometimes called the "closed neighborhood" of Y, and denoted $N[Y]$, while $N(Y) = Y.\eta \backslash Y$ is called the "open neighborhood" [1, 6]). Finally, since all operators map sets into sets, even when we are talking about the neighborhood of a single node, for example z in (1) below, we express it as $\{z\}.\eta$. A **neighborhood closure** operator, φ_η, on $\mathcal{N}$ can be defined by

$$Y.\varphi_\eta = \{z \in Y.\eta \mid \{z\}.\eta \subseteq Y.\eta\}. \quad (1)$$

Readily, $Y \subseteq Y.\varphi_\eta \subseteq Y.\eta$, so φ_η is expansive. It is not hard to see that φ_η is monotone. Finally, since $Y.\varphi_\eta \subseteq Y.\eta$, $Y.\varphi_\eta$ must be idempotent, implying φ_η is a closure operator.[3] The neighborhood closure operator, φ_η, will be fundamental to the development of following sections.

[3] C3, or idempotency, is normally the most difficult property to prove when establishing a closure, or interior, operator.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2021 doi:10.20944/preprints202103.0715.v1

2.1 The Network Interior

Consider any node $y \in N$, and suppose there exists $z \in \{y\}.\varphi_\eta$ implying $\{z\}.\eta \subseteq \{y\}.\eta$. Such a node, z whose "horizon" is contained in that of y, contributes very little to the information content of the network so that its removal from $\{y\}.\eta$ will result in little information loss. This node $z \in \{y\}.\varphi_\eta$ can be reduced. The node y is **irreducible** if $\{y\}.\varphi_\eta = \{y\}$. A sub-network, $\mathcal{I} \subseteq \mathcal{N}$, of irreducible nodes is called the network's **interior**. In the remainder of this section we define an operator, ω, which reduces any network to its irreducible core, and prove that it is almost an interor operator.

If $\{y\}$ is not closed, only elements z in $\{y\}.\eta$ could possibly be in $\{y\}.\varphi_\eta$ so only those need be considered. If $\{z\}.\eta \subseteq \{y\}.\eta$ so that $z \in \{y\}.\varphi_\eta$, we say z is subsumed by y, or z **belongs** to y. We can remove z from N, together with all its connections, and add z to $\{y\}.\beta$, the set of all nodes belonging to $\{y\}$. This set $\{y\}.\beta$ is called its β**-set**. Of course, $y \in \{y\}.\beta$. The cardinality $|\{y\}.\beta|$ is called its β**-count**.

The pseudocode `reduce` of Figure 1 has been used to implement a process ω that reduces any network $\mathcal{N}$ to its irreducible core, $\mathcal{I} = \mathcal{N}.\omega$.

```
while there exist reduceable nodes {
   reducible = 0
   for_each {y} in N {
      for_each {z} in {y}.nbhd - {y} {
         if ({z}.nbhd contained_in {y}.nbhd {
                               //  z is subsumed by y
            remove z from network;
            {y}.beta = {y}.beta union {z}.beta
            reducible = 1 } } } }
```

Figure 1: Reduction code, implementing ω

Applied to $\mathcal{N}_1$, the well-known "karate" network [24], this reduction code yields the interior depicted by bolder links in Figure 2. In this figure, two nodes of the interior have been suffixed by $:n$ to denote their β-count. Only nodes 1 and 33 have non-trivial β-sets of 12 and 8 elements respectively, which have been delimited by dotted lines. (The β-set of node 33 might equally well have been the β-set of node 34; but 33 precedes 34 in the reduction process.)

Proposition 2.1 *The process ω described above is (I1) contractive and (I3) idempotent.*

Proof: Readily ω is contractive and it is idempotent because when $\mathcal{I} = \mathcal{N}.\omega$ is irreducible, the loop is not executed, so $\mathcal{N}.\omega.\omega = \mathcal{I} = \mathcal{N}.\omega$. □

One can show that $\mathcal{N} \subset \mathcal{N}'$ need not imply that $\mathcal{N}.\omega \subset \mathcal{N}'.\omega$, so ω is not an interior operator, even though we call $\mathcal{I} = \mathcal{N}.\omega$ the "interior".

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2021 doi:10.20944/preprints202103.0715.v1

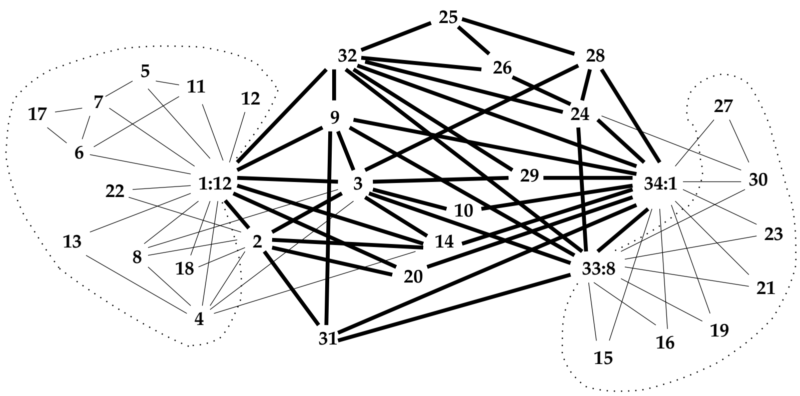

Figure 2: The interior $\mathcal{I}$ of $\mathcal{N}_1$, the Karate network, is shown with bolder links

Proposition 2.2 *Let $\mathcal{I} = \mathcal{N}.\omega$ and $\mathcal{I}' = \mathcal{N}.\omega'$ be irreducible subsets of a finite network $\mathcal{N}$, then $\mathcal{I} \cong \mathcal{I}'$.*

Proof: Let $y_0 \in \mathcal{I}$, $y_0 \notin \mathcal{I}'$. Then y_0 belongs to some point y_1 in $\mathcal{I}'$ and $y_1 \notin \mathcal{I}$ else because $y_0.\eta \subseteq y_1.\eta$ implies $y_0 \in \{y_1\}.\varphi$ so $\mathcal{I}$ would not be irreducible.
Similarly, since $y_1 \in \mathcal{I}'$ and $y_1 \notin \mathcal{I}$, there exists $y_2 \in \mathcal{I}$ such that y_1 belongs to y_2. Now we have two possible cases; either $y_2 = y_0$, or not.
Suppose $y_2 = y_0$ (which is often the case), then $y_0.\eta \subseteq y_1.\eta$ and $y_1.\eta \subseteq y_0.\eta$ or $y_0.\eta = y_1.\eta$. Hence $i(y_0) = y_1$ is part of the desired isometry, i.
Now suppose $y_2 \neq y_0$. There exists $y_3 \neq y_1 \in \mathcal{I}'$ such that $y_2.\eta \subseteq y_3.\eta$, and so forth. Since $\mathcal{I}$ is finite this construction must halt with some y_n. The points $\{y_0, y_1, y_2, \ldots y_n\}$ constitute a complete graph Y_n with $\{y_i\}.\eta = Y_n.\eta$, for $i \in [0, n]$. In any reduction all $y_i \in Y_n$ reduce to a single point. All possibilities lead to mutually isomorphic maps. □

Proposition 2.2 assures us that, even though which nodes are preserved in $\mathcal{I}$ is completely dependent on the order in ω that they are visited, the output must be effectively identical. For example, in Figure 3, assume the nodes x and z are irreducible elements of $\mathcal{I}$. In each case, if $y_0 \in \mathcal{I}$ then, y_1,

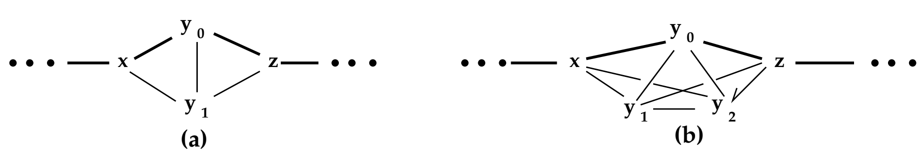

Figure 3: Equivalent nodes y_i in an interior $\mathcal{I}$.

or y_2, could be as well. They are the equivalent nodes defining the isometry. Each set of equivalent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2021 doi:10.20944/preprints202103.0715.v1

nodes must be a “complete” subgraph of $\mathcal{N}$.[4]

A sequence, $\dot{\rho} = \langle y_0, \ldots, y_n \rangle$ of $n+1$ nodes, where $\{y_{i-1}, y_i\} \in L$, or a set of n links $\bar{\rho} = \langle \{y_0, y_1\}, \ldots, \{y_{n-1}, y_n\} \rangle$ is called a **path** $\rho(y_0, y_n)$ of length n.[5] (It is often easier to describe a path in terms of its nodes, $\dot{\rho}$ rather than $\bar{\rho}$ which is more precise.) By $|\rho(x,z)|$ we mean the length of the path independent of whether we are counting nodes or links.

A cycle $\dot{C} = \langle y_0, y_1, \ldots, y_n \rangle$, where $y_n = y_0$, of length $n \geq 4$ is said to have a **bridge** if there exists a path $\bar{\rho}(y_i, y_k) \in L$ where $(k-i)\ mod(n) \neq 1$ [6]. If the path consists of a single link, it is called a **chord**. If C has no such chords it is said to be a **chordless** cycle.[6]

Proposition 2.3 *The nodes of a chordless cycle are irreducible.*

Proof: Let $\{y_{i-1}, y_i\}, \{y_i, y_{i+1}\} \in L$. Suppose $y_i \in y_{i-1}.\varphi_\eta$ implying $y_{i+1} \in \{y_i\}.\eta \subseteq \{y_{i-1}\}.\eta$ or $\{y_{i+1}, y_{i-1}\} \in L$ contradicting chordless assumption. □

Proposition 2.4 *Let $\mathcal{N}$ be a finite network with $\mathcal{I} = \mathcal{N}.\omega$ being an irreducible subset. If $y \in \mathcal{I}$ is not an isolated point then either*

(1) there exists a chordless k-cycle $\dot{C}$, $k \geq 4$ such that $y \in \dot{C}$, or

(2) there exist chordless k-cycles $\dot{C}_1, \dot{C}_2$ each of length ≥ 4 with $x \in \dot{C}_1$ $z \in \dot{C}_2$ and y lies on a path from x to z.

Proof: Let $y_1 \in \mathcal{I}$. Since y_1 is not isolated, let $y_0 \in y_1.\eta$, so $\{y_0, y_1\} \in L$. ≥ 4. Since y_1 is not subsumed by y_0, $\exists y_2 \in y_1.\eta, y_2 \notin y_0.\eta$, and since y_2 is not subsumed by y_1, $\exists y_3 \in y_2.\eta$, $y_3 \notin y_1.\eta$. Since $y_2 \notin y_0.\eta$, $y_3 \neq y_0$.
Suppose $y_3 \in y_0.\eta$, then $\langle y_0, y_1, y_2, y_3, y_0 \rangle$ constitutes a k-cycle $k \geq 4$, and we are done.
Suppose $y_3 \notin y_0.\eta$. We repeat the same path extension. $y_3.\eta \not\subseteq y_2.\eta$ implies $\exists y_4 \in y_3.\eta$, $y_4 \notin y_2.\eta$. If $y_4 \in y_0.\eta$ or $y_4 \in y_1.\eta$, we have the desired cycle. If not $\exists\ y_5, \ldots$ and so forth. Because $\mathcal{N}$ is finite, this path extension must terminate with $y_k \in y_i.\eta$, where $0 \leq i \leq n-3$, $n = |N|$.
The preceding establishes that any link sequence in $\mathcal{I}$ terminates in a cycle of length ≥ 4. Since $\mathcal{N}$ is symmetric, the link sequence could be extended in the opposite direction yielding (2).
So if (1) is not the case, (2) must be. □

The condition that y not be an isolated point is significant. Any tree structured network reduces to a single point, as do many networks consisting of triangles.

Corollary 2.5 *$\mathcal{N}$ is connected if and only if $\mathcal{I}$ is connected.*

A collection of chordless cycles constitutes a cycle system which is itself a matroid [23] with a well defined **rank** [20]. If the network is projected onto a planar representation, then counting those cycles without a bridge yields the rank.

---

[4]A graph, or network, K is said to be complete if for all $x, y \in K$, there is a link $\{x, y\}$. A complete graph on n nodes is denoted by K_n.

[5]Because $\mathcal{N}$ is unidrected, some authors would call these a “walk” [1, 6].

[6]Graphs in which every cycle of length ≥ 4 must have a chord are called “chordal graphs” [9].

 doi:10.20944/preprints202103.0715.v1

Figure 4 illustrates the interior of a small network on 21 nodes. It is a cycle system of rank 5. Here, the links of the interior have been made bolder and again its nodes have their β-counts appended. The β-sets, such as $\{e{:}2\}.\beta$, are suggested by dotted lines. Note that this process effec-

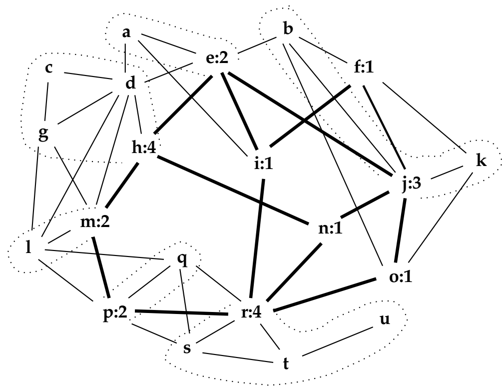

Figure 4: A small network, $\mathcal{N}_2$, of 21 nodes. Interior links are bolder. β-sets are dotted.

tively resolves the question of partitioning networks into disjoint communities [4, 12, 14], without having to specify the number of communities in advance.

2.2 Reduction Performance

Technically, the ω process of Figure 1 is $O(n^2)$ since it can achieve a worst case performance on the unbalanced network of Figure 5 provided the outer loop of the ω code of Figure 1 encounters

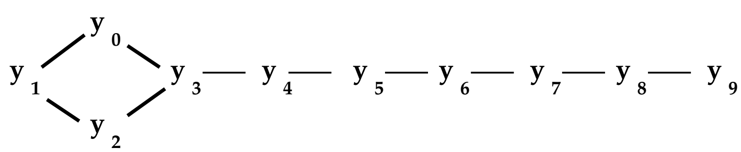

Figure 5: An unbalanced network.

the nodes in order of their subscripts. Then it will remove only one node on each iteration. But in practice, ω appears to actually offer sub-linear performance. With networks of several thousand nodes, ω has never required more than 7 iterations. For example, given the well-known Newman co-authorship network [11] of 363 persons with 823 connecting links, 3 iterations of the outer loop of the ω code of Figure 1 reduces the network to 65 individuals with 111 links constituting its interior shown in Figure 6. (A 4th iteration is required to verify that there are no more reducible nodes.) The node *Stauffer*, in the upper left, has a β-set of 23 elements for which it may be regarded as a surrogate; and the lower left node *Barabasi* has a β-set of 41 elements! In the case of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2021 doi:10.20944/preprints202103.0715.v1

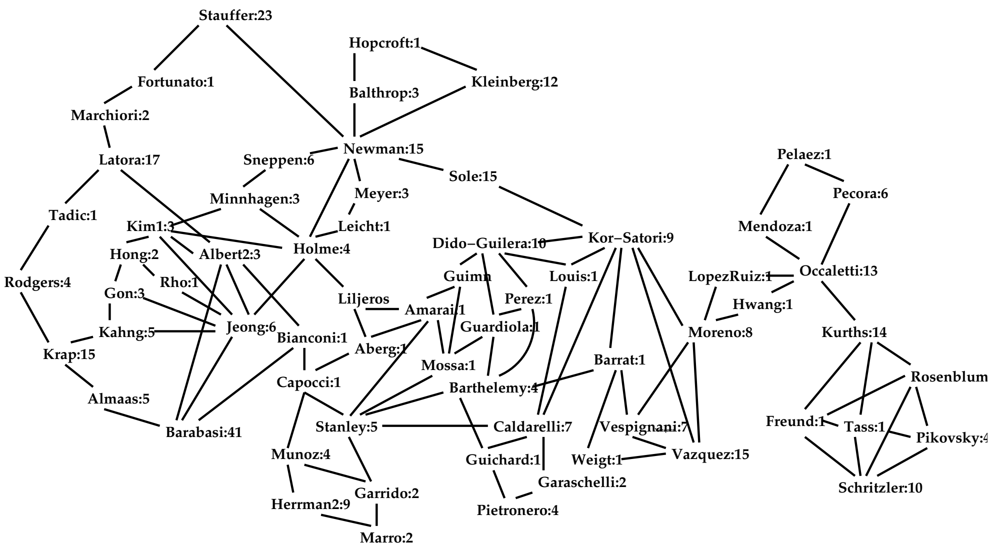

Figure 6: The interior $\mathcal{I}$ of $\mathcal{N}_3$, the 363 node co-authorship network of Newman[11].

Newman co-authorship network, the interior represents a significant reduction in the complexity of the network,

3 Network Properties

There are a number of scalar properties associated with every network $\mathcal{N}$, they include: $n\_nodes = |N|$, $n\_links = |L|$ and $density = |L|/|N|$. The average node degree over all nodes is $2 \cdot density$, since every link has two end nodes [1, 6]. These are trivial to calculate given N and L.

The number of triangles [22] embedded in $\mathcal{N}$ can be calculated by the `count_triangles` whose code is given in Figure 7. Here, the k_count of a link denotes the number of triangles for

```
k_total = 0
for_each link {x, z} in L {
    MEET = {x}.nbhd meet {z}.nbhd
    {x, z}.k_count = cardinality_of(MEET)
    k_total = k_total + {x, z}.k_count }
n_triangles = k_total/3
```

Figure 7: Code counting network triangles

which the link $\{x, z\}$ is one "side". Since that triangle has 3 links, $n\_triangles = k\_total/3$. The computational cost of $\{x\}.nbhd \cap \{z\}.nbhd$ is essentially constant, so the cost of `count_triangles`

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2021 doi:10.20944/preprints202103.0715.v1

is linear, or $O(L)$.

Other scalar properties are dependent on the concept of shortest paths. Let x, z be two nodes in a connected network $\mathcal{N}$. Because $\mathcal{N}$ is connected, there exists a path $\rho(x, z)$ of length n. This may, or may not, be the shortest path (of minimal length) between them. We let $\sigma(x, z)$ denote the (or all) **shortest path**(s) between x and z. The path length $|\sigma(x, z)|$ is also known as the **distance**, $d(x, z)$, between x and y [1, 6]. The $diameter(\mathcal{N})$ of the network is the maximal distance, $d(x, z)$ for all $x, z \in N$. The $eccentricity$ of a node x is $e(x) = max(d(x, z))$ for all $z \in N$. The $radius$, $r(\mathcal{N})$, of the network is minimum eccentricity of any node y [6].

3.1 Communities

Many networks, especially those that represent social connections, are spotted with "clusters" of more densely connected nodes. These clusters of triangular links, which are often called **communities**, arise from the social phenomenon called **triadic closure** [10]. It is known that in many social contexts, if x is connected to y and y is connected to z then x is likely to be connected to z. Even though triadic closure is not really a closure operator[7], its principle has been identified on many repeated occasions [5, 12].

However, we know of no formal definition as to what really constitutes a "community".

There have been numerous efforts to identify communities in a network. Several work on the principle of "bisection" in which removal of certain links separates the network into n distinct communities [4]. A common problem is that usually n must be designated in advance.

A portion of the network that is dense with triangles may be regarded as a community. A connected sub-network of triangles is called a **k-truss** [8]. A connected subset of triangles could be tree-structured, so it is common to specify that a *k-truss* is a connected collection of links with a $k\_count > 1$, where the k_count of a link $\{x, z\}$ is $|\{x\}.\eta \cap \{z\}.\eta|$ as in Figure 7. If $k\_count = 2$, the Karate network of Figure 2 has just one 2-truss, consisting of links connecting the nodes {1, 2, 3, 4, 5, 6, 7, 8, 9, 11, 14, 30, 31, 33, 34} or just less than half the network. It has two 3-trusses connecting the nodes {1, 2, 3, 4, 8, 14} and {9, 24, 33, 34}. The small network of Figure 4 has two 2-trusses of links connecting the nodes {a, e, d, g, h, l, m, p, q, r, s} and {b, f, j, k, o} and four small 3-trusses, which are {b, j}, {d, g, l, m}, {p, q} and {r, s}. There are 23 2-trusses in the Newman network and each is large; but there are only three 8-trusses. They are {Arenas, Dido-Guilera}, {Mano, Occaletti}, and {Barabasi, Jeong, Oltavi, Raven, Schubert}.[8]

The larger values of the principle eigenvector of $A_{n \times n}$ (the adjacency matrix of the network) can indicate well-connected nodes, and often communities [14]. Nodes 1, 3, 33 and 34 of $\mathcal{N}_1$, the Karate network of Figure 2, dominate its principle eigenvector. The principle eigenvector of $\mathcal{N}_2$, the small network of Figure 4, are given in Table 2. Here nodes d, e, m, r stand out. Higher values

[7] As normally encountered, triadic closure is not idempotent. Applied literally, the triadic closure of any network would be the complete graph/network on its n nodes.

[8] Arenas, Mano, Oltavi, Raven, Schubert are not elements of the interior shown in Figure 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2021 doi:10.20944/preprints202103.0715.v1

in this eigenvector appear to correlate with higher node degree. The nodes $Barabasi$, $Jeong$, and $Oltvai$ (in $\{Jeong\}.\beta$) are most prominant in the eigenvector of the Newman network.

All of these methods have been proposed to denote "communities". We would suggest that the β-sets attached to $\mathcal{I}$ also denote "communities".

3.2 Important Nodes

A fundamental quest in the analysis of many networks is the identification of its "important" nodes. They may be a node of high degree in a community, but need not be. In social networks, "importance" may also be defined with respect to the path structure [3, 13]. Those nodes, $C_d = \{y \in \mathcal{N}\}$ for which the eccentricity, $e(y)$, or $\sum_{x \neq y} d(x, y)$, is $minimal$, have traditionally been called the **center** of $\mathcal{N}$ [1, 6], they are "closest" to all other nodes. It is well known that this subset of nodes must be edge connected. One may assume that these nodes in the "center" of a network are "important" nodes.

Alternatively, one may consider those nodes which "connect" many other nodes, or clusters of nodes, to be the "important" ones. Let $nsp_{xz}(y)$ denote the *number* of shortest paths $\sigma(x, z)$ containing y; then those nodes y for which $nsp_{xz}(y)$ is $maximal$ are those nodes that are involved in the most connections. Let $C_b = \{y \in \mathcal{N}\}$, for which $nsp_{xz}(y)$ is *locally* maximal. This is sometime called "betweenness centrality" [2, 3, 13].[9]

3.3 Network Properties Preserved by the Interior

The next 3 lemmas, culminating in Proposition 3.4 help clarify the interaction of β-sets with the nodes of $\mathcal{I}$. In these lemmas, we assume that x_0, y_0 and $z_0 \in \mathcal{I}$.

Lemma 3.1 *Let $y_k \in \{y_0\}.\beta$. There exists a node sequence $\langle y_0, y_1, \ldots, y_k\rangle$ such that $y_i \in \{y_0\}.\beta$, $0 \le i \le k$.*

Proof: In the reduction process of Figure 1, if y_{i+1} is subsumed by y_i, then $\{y_{i+1}\}.\beta \subset \{y_i\}.\beta$ yielding the chain of nested sets $\{y_k\}.\beta \subset \{y_{k-1}\}.\beta \subset \ldots \subset \{y_0\}.\beta$. □

Note that even if $y_i \in \{y_{i-1}\}.\eta$ belongs to y_{i-1}, there may be other nodes $x_i \in \{y_{i-1}\}.\eta$ such that $x_i \notin \{y_{i-1}\}.\beta$.

Lemma 3.2 *Let $\langle y_0, \ldots, y_k\rangle \in \{y_0\}.\beta$ and let $\{y_k, z\} \in L$ where $z \notin \{y_0\}.\beta$. Then for all $y_i, 0 \le i \le k$, $\{y_i, z\} \in L$.*

Proof: By the reduction process ω, when y_i is subsumed by y_{i-1}, $\{y_i\}.\eta \subseteq \{y_{i-1}\}.\eta$. So if $\{y_i, z\} \in L$ then $z \in \{y_{i-1}\}.\eta$ or $\{y_{i-1}, z\} \in L$. □

[9] In [13], Newman proposes the notion of "random walk betweenness" as an alternative to shortest path betweenness.

 doi:10.20944/preprints202103.0715.v1

Lemma 3.3 *Let $x \in \{x_0\}.\beta, z \in \{z_0\}.\beta$ where $x_0, z_0 \in \mathcal{I}$. If $\{x, z\} \in L$, then there exists $y \in \mathcal{I}$ such that $\{x, y\}, \{y, z\} \in L$.*

Proof: By Lemma 3.2 we know $\exists \{x, z_0\}, \{z, x_0\} \in L$. If $\{x_0, y_0\} \in L$ we are done. So suppose not. By Prop. 2.4 we can assume $\exists y \in \mathcal{I}$ (or a sequence y_i) such that $\{x_0, y\}, \{y, z_0\} \in L$. We claim $\{x, y\} \in L$, since otherwise $\langle y, x_0, \ldots, x, \ldots, z_0, y\rangle$ is a chordless cycle of length ≥ 4, and hence by Prop. 2.3 is irreducible. Similarly $\{y, z_0\} \in L$. □

Two β-sets, $\{x_0\}.\beta, \{y_0\}.\beta$ are said to be β-**connected** if there exists $x, y \neq x_0, y_0$ where $x \in \{x_0\}.\beta$, $y \in \{y_0\}.\beta$ and $\{x, y\} \in L$. The preceding lemmas describe links that must exist if β-sets are connected. These are illustrated in Figure 8. In this figure, solid lines denote links

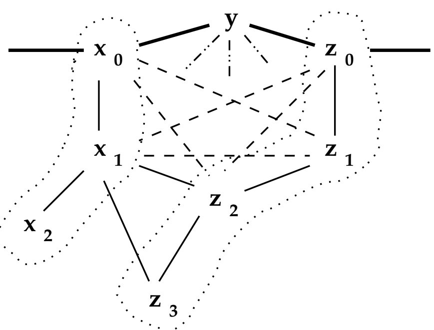

Figure 8: Links that can be inferred between connected β-sets.

that are "known" to exist for one reason or another. The dotted (. . .) lines that enclose β-sets were established by the reduction process. Each conforms to Lemma 3.1. Observe that the entire set of nodes, $\{x_1, x_2, z_1, z_2, z_3\}$ could constitute either $\{x_0\}.\beta$ or $\{z_0\}.\beta$ depending solely on the order of node reduction. This has been illustrated in $\mathcal{N}_1$, Figure 2, where $\{33\}.\beta$ could have been $\{34\}.\beta$. Proposition 2.2 establishes that the structure of the interior, $\mathcal{I}$, is independent of the order in which nodes are encountered in the ω process; but the structure of β-sets produced by the code `reduce` can be very dependent on this order.

The dashed links (- - -) denote links that can be inferred from Lemma 3.2. For instance, z_1 can not subsume z_2 unless $x_1 \in \{z_1\}.\eta$ because $x_1 \in \{z_2\}.\eta$. The (- ·· -) links connecting y to the nodes x_1, z_1, z_2 can be inferred from Proposition 2.3.

While in many networks the β-sets will be separated (as in Figure 4), they may be links between them. It is not hard to imagine a link between $a \in \{e\}.\beta$ and $c \in \{h\}.\beta$. The lemmas establish that either such a link must introduce a new chordless cycle into $\mathcal{I}$, or else there must be an abundance of "triangles" surrounding the network interior.

Proposition 3.4 *Let $\rho(x, z)$ be a path where $\{x, z\} \notin L$ (i.e. $|\rho(x, z)| \geq 2$) and $x \in \{x_0\}.\beta$ and $z \notin \{x_0\}.\beta$. Then there exists a path $\rho'(x, y, z)$ where $y \in \mathcal{I}$ and $|\rho'(x, y, z)| \leq |\rho(x, z)|$.*

Proof: We may assume that $\rho(x, z) \cup \mathcal{I} = \emptyset$, else there is nothing to prove. So we may also assume that $\rho(x, z)$ lies entirely within connected β-sets. By Lemma 3.3, $\exists y \in \mathcal{I}$ such that $\{x, y\}, \{y, z\} \in L$, so $|\rho'(x, y, z)| = 2 \leq |\rho(x, z)|$. □

3.4 Network Centrality

Proposition 3.5 *If $\mathcal{N}$ is not unbalanced then the center C_d (in terms of distance) is an element of (or intersects with) the interior $\mathcal{I}$ of $\mathcal{N}$.*

Proof: If x and z are in separated β-sets then $\sigma(x, z) = \langle x = x_k, x_{k-1}, \ldots, x_0\rangle \cup \langle y_1, \ldots, y_m\rangle \cup \langle z_0, \ldots, z_n\rangle$ where $y_1 = x_0$, $y_m = z_0$ and $y_i \in \mathcal{I}$. Since $\mathcal{N}$ is not unbalanced, we may assume $k \approx n$, so the center of $\sigma(x, z)$ is one of the $y_1, \ldots, y_m$.
If x and z are in connected β-sets and $|\rho(x, z)| \geq 2$, then Prop. 3.4 establishes the existence of a shortest path through $\mathcal{I}$ as well.
If $x, z \in \{x\}.\beta$ then no shortest path involves $\mathcal{I}$; but since $\mathcal{N}$ is not unbalanced, these constitute a small number of cases and can be ignored. □

In Figure 3(b), if y_1 is in the center C, then so are y_0 and y_2, implying $C \cap I \neq \emptyset$.

Proposition 3.5 requires that $\mathcal{N}$ not be too unbalanced. Figure 5 illustrates why. It is not hard to show that y_5 is the center with maximum distance over all x being $d(x, y_5) = 4$. Our rule of thumb is that a network is reasonably well-balanced if given any $x \in \{x_0\}.\beta$ then the probability that a randomly chosen y is also in $\{x_0\}.\beta$ is small, that is $pr(y \in \{x_0\}.\beta | x \in \{x_0\}.\beta) < \varepsilon$ where $\varepsilon < 0.20$.

Proposition 3.6 *If $\mathcal{N}$ is not unbalanced, then any center C_b of $\mathcal{N}$(in terms of betweenness) is an element of $\mathcal{I}$.*

Proof: This proof follows the line of Prop. 3.5 in which, unless x and z are in the same β-set, all shortest paths $\sigma(x, z)$ either involve $\mathcal{I}$ or have a path $\rho'(x, y, z)$ of the same length through $\mathcal{I}$. Hence a node y for which $\sigma_{x,z}(y)$ is maximal will be an element of $\mathcal{I}$. □

That $\mathcal{I}$ contains the betweenness center is evident in the Karate network of Figure 2 and the small network of Figure 4.

Figure 9 illustrates a somewhat different "unbalanced" network in which x and $z \notin \mathcal{I}$ are betweenness centers. One can calculate that $nsp(x) = nsp(z) = 6 * 6 + 4 * 6 * 6 + 4 * 6 = 204$ which are locally maximal.

Calculating betweenness centers is computationally expensive, even with improved algorithms such as [2]. Knowing that they must exist in the interior $\mathcal{I}$ and restricting the calculation to just those nodes can greatly improve performance, especially when betweenness is employed in other

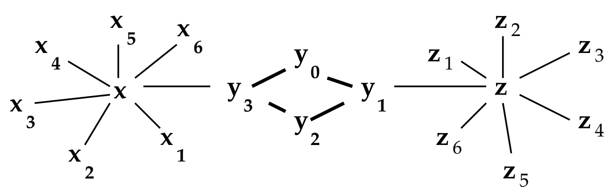

Figure 9: Another unbalanced network.

procedures such as [4]. Consequently, dwelling too much on unbalanced networks can be self defeating since the majority, and possibly almost all, networks are well-balanced.

4 Network Generation by Expansion

The interior, $\mathcal{I}$, of a network $\mathcal{N}$ represents its global structure. If the β-count is appended to each node of $\mathcal{I}$, how well does $\mathcal{I}$ represent $\mathcal{N}$ as a whole? In effect, what is the information content of $\mathcal{I}$, so augumented?

One measure of the information content of any collection of network properities is the ability to construct, or generate, similar networks based on those properties. For example, given a network $\mathcal{N} = (N, L)$ one can construct many different networks $\mathcal{N}' = (N', L')$ such that $|N'| = |N|$ and $|L'| = |L|$. But they need not be at all similar to $\mathcal{N}$. Here we are using "similar" in its colloquial sense. A formal notion of "similarity" would require it to be an equivalence relation.[10] One way of determining the nature of networks with a given interior, $\mathcal{I}$, and known β-counts is to randomly generate some. Let $\mathcal{I}$ be given. Suppose the β-count of a node y is greater than one. New nodes can be attached to replace those of the original β-set. Let $y{:}n$ be the node to be expanded ($n > 1$), and let z denote the new node.[11] Besides the link $\{y, z\}$, we require $\{z\}.\eta \subseteq \{y\}.\eta$. A random number determines how many of the other nodes in $\{y\}.\eta$ will be linked to z, and which, if any, of those are also randomly chosen.

In the reduction process, ω, nodes with considerable β-sets may be subsequently reduced themselves. In the re-expansion, a portion of the β-count of y may be transferred to the β-count of z. Pseudocode for a procedure `expand` to implement an operator ε that generates new nodes relative to the interior is given in Figure 10.[12]

As a test, the interior $\mathcal{I}$ of $\mathcal{N}_2$, Figure 4, has been expanded 3 times (using different random number seeds) to yield $exp.1$, $exp.2$ and $exp.3$ of Figure 11.

[10] It has been proposed that a formal definition of "similarity" must require the similar networks to have isomorphic interiors [19].

[11] Our code generates artificial node names of the form 'A0, B0, ..., Z0, A1, ...'. The last generated node in the expansion of Figure 6 is $M11$.

[12] ε, as shown here is a round-robin procedure expanding one node in a β-set at a time. An alternative, and slightly faster, process can be found in [18].

 doi:10.20944/preprints202103.0715.v1

```
while still_expanding {
    still_expanding = 0
    for_each  y  in NODES {
        if (y.beta_count > 1)
            { z = new_node()
            add new_node to NODES
            chosen = choose_subset (y.nbhd)
                        // distribute some of y.beta_count to z
            increment = y.beta_count/(n_chosen+1)
            y.beta_count = y.beta_count - increment
            z.beta_count = 1 + increment
            add (y, z) to LINKS
                        // link z to chosen nodes in y.nbhd
            for_each  x  in chosen {
                add (x, z) to LINKS }
            still_expanding = 1 } } }
```

Figure 10: Pseudocode for a procedure which generates similar networks.

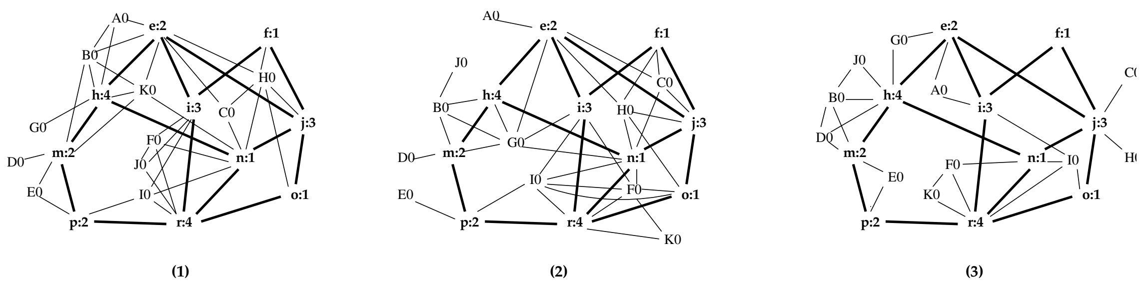

Figure 11: Three different expansions of $\mathcal{I} = \mathcal{N}_2.\omega$, Figure 4.

Proposition 4.1 *Let $\mathcal{I}$ be the interior of a network $\mathcal{N}$, that is $\mathcal{I} = \mathcal{N}.\omega$, then $\mathcal{I}.\varepsilon.\omega = \mathcal{I}$.*

Proof: The expansion procedure of Figure 10 was written to make this true. Consider z_n, the last node appended by ε. By construction, $x_n.\eta \subseteq y_n.\eta$ for some y_n; so z_n can be subsumed into $y_n.\beta$. A finite induction $z_n, \dots, z_1$ completes the proof. □

Considered as operators, ε is a left-inverse of ω since $\varepsilon \cdot \omega = 1$, where 1 denotes the identity operator. However, $\omega \cdot \varepsilon \neq 1$, as shown by Figure 11.

To what extent are the network features of $\mathcal{N}$ enumerated in the preceding section preserved in the randomly generated networks, $\mathcal{N}.\omega.\varepsilon$? Readily, the generation process ε has been constrained so that $|\mathcal{N}.\omega.\varepsilon| = |\mathcal{N}|$ and $\mathcal{N}.\omega.\varepsilon.\omega = \mathcal{I} = \mathcal{N}.\omega$, so path based centers of Section 3.4 are preserved. Some other network properties are illustrated in Table 1.

Table 2 presents the primary eigenvector associated with the nodes of $\mathcal{N}_2$ in Figure 4 and for the three expansions shown in Figure 10. (Note that, except for the ten nodes of $\mathcal{I}$, node values for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 March 2021 doi:10.20944/preprints202103.0715.v1

	$\|N\|$	$\|L\|$	$density$	$triangles$	$2\_trusses$	$3\_trusses$
$\mathcal{N}_2$	21	44	2.095	21	2	4
exp.1	21	49	2.333	31	1	3
exp.2	21	46	2.190	25	2	3
exp.3	21	37	1.762	13	2	2

Table 1: Network properties of networks in Figure 11 generated from $\mathcal{I} = \mathcal{N}_2.\omega$, Figure 4.

generated expansions are not comparable with node values of the original $\mathcal{N}$.)

	a	b	c	d	e	f	g	h	i	j	k
$\mathcal{N}$	0.179	0.182	0.123	0.350	0.293	0.155	0.226	0.234	0.194	0.231	0.120
	A0	B0	C0	D0	e	f	E0	h	i	j	F0
exp.1	0.170	0.295	0.225	0.033	0.355	0.129	0.053	0.306	0.202	0.265	0.162
exp.2	0.048	0.095	0.203	0.021	0.262	0.183	0.026	0.192	0.254	0.285	0.303
exp.3	0.125	0.212	0.056	0.187	0.265	0.120	0.093	0.353	0.270	0.243	0.195
	l	m	n	o	p	q	r	s	t	u	
$\mathcal{N}$	0.291	0.293	0.159	0.174	0.271	0.220	0.280	0.187	0.104	0.022	
	G0	m	n	o	p	H0	r	I0	J0	K0	
exp.1	0.054	0.190	0.387	0.133	0.112	0.265	0.224	0.164	0.104	0.272	
exp.2	0.192	0.118	0.379	0.271	0.142	0.253	0.325	0.307	0.017	0.115	
exp.3	0.144	0.236	0.336	0.208	0.163	0.056	0.369	0.276	0.132	0.132	

Table 2: Value of nodes in Figures 4 and 10 as expressed by the primary eigenvector.

This section began with the question "how well does $\mathcal{I}$ represent $\mathcal{N}$ as a whole?" Figure 11 and Tables 1 and 2 provide abundant evidence that given just $\mathcal{I}$, with each node augmented with its β-count, a random process can generate new networks whose properties are very similar to those of $\mathcal{N}$. It would seem to be a very good description of $\mathcal{N}$.

5 Observations

This paper might have been titled "An Operator Approach to ..." since the operators η, φ_η, ω and ε play such an important role. This aspect is briefly suggested by Proposition 4.1, but not enlarged. But surely, interesting networks are dynamic; they change over time which demands an operator approach. So one might ask: "is a transformation $\tau : \mathcal{N} \to \mathcal{N}'$ continuous?" [17]. The operators ω and ε are, in fact, "continuous" with respect to φ_η. Moreover, it appears that $\mathcal{N}.\nu = \mathcal{N} \backslash \mathcal{N}.\omega = \mathcal{N} \backslash \mathcal{I}$ is a violator space in the sense of [7]. There could be a lot more here.

However, computability is such a dominant theme in current network analysis and understanding that we thought focusing on the use of set-theoretic computer procedures such as `reduce`, `count_triangles` and `expand` was more important. Programming with set operators is not wide spread. Yet, these set-theoretic procedures appear to be quite scalable. (Execution of none exceeded 1 second (the smallest unit on the timer we are using); only the calculation of the eigenvectors of Figure 6 exceeded 5 seconds.)

Only standard set-theoretic reasoning has been used to develop the reduction process, ω, that leads to the concept of the "interior", $\mathcal{I}$, of a network, $\mathcal{N}$, and its β-set. It is a powerful concept that effectively captures the essence of many networks, as shown by Section 4 in which very similar networks can be generated from $\mathcal{I}$ alone. Moreover, by reducing a network to its interior, one effectively partitions the network into it constituent β-set communities.

But the reduction process has its limitations. Some networks are nearly irreducible to start with. The sparse network of Norwegian corporate directors [21] is an example. And hierarchical networks reduce to a single node.[13] Other networks can be too dense. The complete network K_n also reduces to a single node. Still, we believe that the easily computed interior is a most effective network descriptor and possibly should be an automatic first step in network description and understanding.

Acknowledgment: The author would thank Christopher Taylor for coding the set manipulating C^{++} code, and John Burkardt who made his eigenvector code available on the internet.

References

[1] Geir Agnarsson and Raymond Greenlaw. *Graph Theory: Modeling, Applications and Algorithms*. Prentice Hall, Upper Saddle River, NJ, 2007.

[2] Ulrik Brandes. A Faster Algorithm for Betweeness Centrality. *J.Mathematical Sociology*, 25(2):163–177, 2001.

[3] Linton C. Freeman. Centrality in Social Networks, Conceptual Clarification. *Social Networks*, 1:215–239, 1978/79.

[4] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.*, 99:7821–7826, 2002. doi:/10.1073/pnas.122653799.

[5] Mark S. Granovetter. The Strength of Weak Ties. *Amer. J. of Sociology*, 78(6):1360–1380, 1973.

[6] Frank Harary. *Graph Theory*. Addison-Wesley, 1969.

[7] Yulia Kempner and Vadim E. Levit. Violator spaces vs closure spaces. *European Journal of Combinatorics*, 80:203–213, 2019. doi:10.1016/j.ejc.2018.07.012.

[8] Ian McCulloh and Onur Savas. k-Truss Network Community Detection. In *2020 IEEE/ACM Intern. Conf. on Advances in Social Networks Analysis and Mining (ASONAM)*, The Hague, Netherlands, 2020. (virtual via Zoom).

[9] Terry A. McKee. How Chordal Graphs Work. *Bulletin of the ICA*, 9:27–39, 1993.

[13] A single node interior with a very large β-set.

 doi:10.20944/preprints202103.0715.v1

[10] Gerald Mollenhorst, Beate Völker, and Henk Flap. Shared contexts and triadic closure in core discussion networks. *Social Networks*, 34:292–302, Jan. 2012.

[11] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[12] M. E. J. Newman. Detecting community structure in networks. *European Physical J. B*, 38:321–330, 2004.

[13] M. E. J. Newman. A measure of betweenness centrality based on random walks. *Social Networks*, 27:39–45, 2005.

[14] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys.Rev.E*, 74(036104):1–22, July 2006.

[15] Mark Newman. *Networks*. Oxford Univ. Press, Oxford, U.K., 2018. doi:101093/0s0/9780198805090.001.0001.

[16] Ratko Orlandic, John L. Pfaltz, and Christopher Taylor. A Functional Database Representation of Large Sets of Objects. In Hua Wang and Mohamed A. Saraf, editors, *25th Australasian Database Conference, ADC 2014*, pages 189–196, Brisbane, QLD, 2014.

[17] John Pfaltz and Josef Šlapal. Transformations of discrete closure systems. *Acta Math. Hungar.*, 138(4):386–405, 2013.

[18] John L. Pfaltz. Computational Processes that Appear to Model Human Memory. In *Proceedings 4th International Conference, Algorithms for Computational Biology (AlCoB 2017)*, pages 85–99, Aveiro Portugal, 2017.

[19] John L. Pfaltz. Graph Similarity Defined by Graph Transformation. In Imre Rudas, Lambros Ekonomou, and George Vachtsevanos, editors, *2nd International Conf. on Applied Math and Computational Science*, page 35 (abstract), Budapest, Hungary, 2018.

[20] John L. Pfaltz. Cycle Systems. *Math. for Applications*, 9:55–66, 2020. doi:1013164/ma2020.5.

[21] C. Seierstad and Tore Opsahl. For the few not the many? The effects of affirmative action on presence, prominence, and social capital of female directors in Norway. *Scandinavian J. of Management*, 27:44–54.

[22] Charalampos E. Tsourakakis, Petros Drineas, Eirinaios Michelakis, Ioannis Koutis, and Christos Faloutos. Spectral counting of triangles via element-wise sparsification and triangle-based link recommendation. *Soc. Network Analysis and Mining*, 1(2):75–81, Apr. 2011.

[23] Neil White. *Theory of Matroids*. Cambridge University Press, Cambridge, UK, 1986.

[24] Wayne W. Zachary. An Information Flow Model for Conflict and Fission in Small groups. *J. of Anthropological Research*, 33(4):452–473, 1977.