Graph Similarity
Defined by Graph Transformation

John L. Pfaltz

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

Abstract

Similarity is an important way of categorizing mathematical objects. In this paper we consider how similarity between undirected graphs, or networks, can be defined. We do this by reducing any undirected graph, $G = (V, E)$ to its *irreducible spine*, I. First we show that the reduction process, ω, is a unique transformation and thus is a well-defined function. Each irreducible spine, I, thus defines an equivalence class consisting of all graphs $\{G_i\}$ such that $G_i.\omega = I$.

Reduction by ω preserves many key properties. Specifically it preserves shortest path lengths in I, and (under rather general conditions) preserves the location of distance and betweeness centers. In essence, every member graph of the equivalence class will have much the same path structure.

Next, an algorithm ε is presented which, given any irreducible spine I, will generate (randomly) a graph G_k in the equivalence class defined by I, that is, we must have $G_k.\omega = I$. Since $I.\varepsilon.\omega = I$, $\varepsilon = \omega^{-1}$.

It is of some importance in network analysis to be able to quantify the notion of network similarity; to be able to say that a network G_1 is "closer" to the network G_2 than to G_3. A popular method of comparing similarity between two networks G_1 and G_2 is to calculate the Spearman coefficient based on corresponding node degrees. We do this for 5 random graphs in the equivalence class of our running example.

Finally, we show that any irreducible spine consists of a system of cycles, which we show to be a matroid. We then note that every spine, I, is topologically equivalent to a cubic graph. Thus individual cubic graphs generate a second, and coarser, equivalence relation on the family of all undirected graphs, which we call its genus.

Key words: Equivalence, Cubic graph, Closure, Undirected Graph, Irreducible

1. Introduction

How similar is a graph G_1 to a graph G_2? This can be an important question in many applied fields, including chemical structure analysis, data mining algorithms, social network investigation and cognitive studies. Consequently, there is no shortage of methods to measure graph "similarity"; the collection [12, 19, 30, 31] provides a small representative sample. All are statistical in nature and yield a "similarity value" in the range [0.0, 1.0]. Many are variations on the Spearman rank correlation measure which we will explore in more detail in Section 3. These similarity measures may have considerable value in applications, but they are not examples of a mathematical "similarity" relation.

Let $G = (V, E)$ denote a **graph** on a set V of elements, or vertices, with an edge set E. For this paper we assume that the graph is undirected, that is $(x, y) \in E$ implies $(y, x) \in E$ (or alternatively, E is a collection of 2-element sets $\{x, y\}$).[1] Mathematical similarity is an **equivalence** relation. That is, it is *reflexive* (every graph G is similar to itself), *symmetric* (if G_1 is similar to G_2, then G_2 must be similar to G_1) and *transitive* (if G_1 is similar to G_2 and G_2 is similar to G_3 then G_1 must be similar to G_3 as well) [1, 27]. Any equivalence relation, say $\sim$, partitions its domain into disjoint **equivalence classes** and such a partition defines an equivalence. Two very simple equivalence relations are $\sim_1$ and $\sim_2$, where $G \sim_1 G'$ if $|V| = |V'|$ and $G \sim_2 G'$ if $|E| = |E'|$. Equivalence relations are closed under conjunction, that is $\sim_1 \wedge \sim_2 = \sim_3$.

We will define a concept of similarity by means of graph transformation. Readily, any functional transformation induces a similarity relation, by letting $X \sim Y$ if $f(X) = f(Y)$. Then the inverse image sets $f^{-1}(f(X))$ become the equivalence classes. The problem lies in establishing that the graph transformation is single-valued, and that its equivalence classes preserve some property of interest.

1.1. Neighborhoods and Closure

By the neighborhood of a singleton set $\{x\}$, denoted $\{x\}.\eta$, we mean the set of all $y \in V$ such that $(x, y) \in E$. We assume that $x \in \{x\}.\eta$, so η is a reflexive

[1] Some of our results here can be formulated when the graphs are directed; but both their statement and their proof tend to be messy.

(or expansive) operator.[2] By the **neighborhood** of a set X, we mean

$$X.\eta = \cup_{x \in X} \{x\}.\eta \tag{1}$$

We denote operators such as η, which map sets into sets, using a suffix notation, $X.\eta$, to distinguish them from more common single element valued functions, such as $f(x)$.

An operator φ is said to be a **closure** operator if for all X, Y,

C1: $X \subseteq X.\varphi$ (expansive),

C2: $X \subseteq Y$ implies $X.\varphi \subseteq Y.\varphi$ (monotone), and

C3: $X.\varphi.\varphi = X.\varphi$ (idempotent)

Readily, the neighborhood operator, η, is expansive and monotone; but almost never idempotent. Iterated neighborhood operations just keep getting bigger and bigger. A set X is **closed** if $X.\varphi = X$. A graph G is said to be **irreducible** if all its singleton subsets are closed; that is $\{x\}.\varphi = \{x\}$.

An operator that is defined by taking a union of smaller subsets, such as η in (1), is said to be **extensible**. Because η is extensible, it can be represented by the edges in a graph. A closure operator, φ, is not, in general, extensible

Using the neighborhood concept η, we can define the **neighborhood closure** operator[3] φ_η to be

$$Y.\varphi_\eta = \{x | \{x\}.\eta \subseteq Y.\eta\}. \tag{2}$$

Effectively, $Y.\varphi_\eta$ consists of all elements of Y itself, plus those adjacent elements x with no larger a relational horizon than that of Y. In a social network, this is indeed a "closed" group! While there can be many closure operators, such as transitive closure, on an set of elements, we are solely concerned with φ_η in this paper. Readily, for all Y,

$$Y \subseteq Y.\varphi_\eta \subseteq Y.\eta. \tag{3}$$

Proposition 1.1. *If η is an extensible operator, then φ_η is a closure operator.*

Proof: (C1) Readily, $Y \subseteq Y.\varphi_\eta$ by definition.
(C2) Let $X \subseteq Y$ and let $z \in X.\varphi_\eta$. By (2) $\{z\}.\eta \subseteq X.\eta \subseteq Y.\eta$ hence $z \in Y.\varphi_\eta$.
(C3) Let $z \in Y.\varphi_\eta.\varphi_\eta$. Then $\{z\}.\eta \subseteq Y.\varphi_\eta.\eta = \bigcup_{x \in Y.\varphi_\eta} \{x\}.\eta \subseteq Y.\eta$, hence $z \in Y.\varphi_\eta$. □

[2] Some authors [1, 11] call this a *closed* neighborhood. But, we have a very different concept of closure.

[3] Neighborhood closure has been called "experiential" and "relational" closure in other papers [24, 25].

To see why extensibility is required to establish (C3), let $V = \{a, b, c\}$ and let $\{a\}.\eta = \{ab\}$, $\{b\}.\eta = \{ab\}$, $\{c\}.\eta = \{c\}$, but $\{ab\}.\eta = \{abc\}$. η is not extensible and $\{a\}.\varphi_\eta.\varphi_\eta = \{ab\}.\varphi_\eta = \{abc\} \neq \{ab\} = \{a\}.\varphi_\eta$, so φ_η as defined by (2) is not idempotent. (We often elide the commas when denoting a set by enumeration.)

2. Graph Transformations

There are various ways of defining a transformation, f, between two graphs $G = (V, E)$ and $G' = (V', E')$. We denote a graph by its basic set of elements, or vertices, V and some relation on these elements such as an edge set E. A common way is to first define a function, f, that maps V into V' and constrain this function to satisfy certain constraints with respect to the edge sets E and E'. For example, the traditional definition of a *graph homomorphism* requires that if $(x, y) \in E$ then $(f(x), f(y)) \in E'$, *c.f.* [1, 11]. We, however, will employ a rather different transformation concept.

A **transformation**, $G \xrightarrow{f} G'$ between two graphs $G = (V, E)$ and $G' = (V', E')$, is a function f whose domain is 2^V (the powerset of V) and whose codomain is $2^{V'}$. It may be constrained to satisfy conditions on an operator over these subsets of V and V', such as the edge operator E and E', or other operators such as the neighborhood operator η or a closure operator φ. Readily, an operator is just a transformation in which $V' = V$. A transformation is said to be **monotone** if for all $X, Y \neq \emptyset \subseteq V$, if $X \subseteq Y$ then $X.f \subseteq Y.f$.[4] A transformation, $G \xrightarrow{f} G'$, is said to be **continuous** with respect to φ if for all Y, $Y.\varphi.f \subseteq Y.f.\varphi'$ [17, 28].

Proposition 2.1. *If $G \xrightarrow{f} G'$ and $G' \xrightarrow{g} G''$ are monotone, continuous transformations, then $G \xrightarrow{f.g} G''$ is monotone and continuous.*

Proof: See [20].
We will not make much use of this continuity concept in this paper; but it can be valuable when modeling dynamic systems.

There are an number of advantages to using this powerset approach. One is that it supports a clean, functional way of adding and deleting elements in a graph. To delete a subset X of elements, one can let $X.f = \emptyset$. Then by monotonicity,

[4] Requiring X, Y to be non-empty avoids several awkward problems.

for all Y with $X \subset Y$, $Y.f \subseteq (Y \backslash X).f$. Similarly, one can add a subset X' to the elements of V' by the function g in which $\emptyset.g = X'$. Here the condition on monotonicity is important because we don't necessarily want $\emptyset.g = X' \subseteq Y' = Y.g$ for all Y' just because $\emptyset \subseteq Y$ for all Y.

2.1. Graph Reduction, ω

Let G be a graph (V, E), with a neighborhood operator η. Suppose $z \in \{y\}.\varphi_\eta$, implying by (2) that $\{z\}.\eta \subseteq \{y\}.\eta$. Since $\{z\}.\varphi_\eta \subseteq \{y\}.\varphi_\eta$, the set $\{z\}$ contributes nothing to the closure structure of G; it can be removed from G with little loss of information. We define the transformation $\omega_z : G \to G'$ by $\{z\}.\omega_z = \emptyset$ where ω_z is the identity map on $V \backslash \{z\}$ and $\{u', v'\} \in E'$ if and only if $\{u, v\} \in E$, $u, v \neq z$. We say z **belongs** to y denoted by $z \in \{y\}.\beta$.[5] It is not hard to show that ω_z is monotone. and continuous since $z \in \{y\}.\varphi_\eta$.

If ω_z is iterated until there are no vertices y such that $z \in \{y\}.\varphi_\eta$ then G is **irreducible**. We call this irreducible subgraph, I, in which all singleton subsets $\{y\}$ are closed, the irreducible **spine** of G. The following computer procedure, *reduce* implements ω.

```
while there exist reducible nodes
  {
  for_each y in V
     {
     get {y}.nbhd
     for_each z in {y}.nbhd - {y}
        {
        if ({z}.nbhd contained_in {y}.nbhd
           {        //  z belongs to  y
           remove z from network
           add z to {y}.beta
           }
        }
     }
  }
```

(Operators, such as `contained_in` are features of our C++ implementation [18].) It repeatedly sweeps through all vertices $y \in V$, deleting any vertices $z_i \in \{y\}.\varphi_\eta$, together with all edges incident to z_i, until no such z remain in V. That is, $\omega =$

[5] As with η, we also assume $y \in \{y\}.\beta$.

$\omega_{z_1} \cdot \omega_{z_2} \cdot \ldots \cdot \omega_{z_n}$. Since each ω_{z_i} is monotone and continuous, ω is as well, that is $Y.\varphi_\eta.\omega \subseteq Y.\omega.\varphi_\eta{}'$. The process terminates when every singleton subset $\{y\} \subseteq I$ is closed.[6]

Below we show that $G' = G.\omega$ is unique (up to isomorphism) regardless of the order in which the vertices $y \in V$ are visited by ω or the order in which vertices $z \in \{y\}.\varphi_\eta$ are deleted.

Proposition 2.2. *Let $I = G.\omega$ and $I' = G.\omega'$ be two irreducible subsets of a finite graph G, then $I \cong I'$.*

Proof: Suppose $I \neq I'$. Let $y_0 \in I$, $y_0 \notin I'$, so $\exists y_1 \in I$ where $y_0 \in \{y_1\}.\varphi_\eta$, or equivalently, $\{y_0\}.\eta \subseteq \{y_1\}.\eta$. $y_1 \notin I$ because $y_0 \in \{y_1\}.\varphi_\eta$ would contradict the irreducible nature of I. Since $y_1 \notin I$, there must exist $y_2 \in I$ such that $y_1 \in \{y_2\}.\varphi_\eta$.
We claim $y_2 = y_0$, because if not, $y_0 \in \{y_1\}.\varphi_\eta$ implies $\{y_0\}.\eta \subseteq \{y_1\}.\eta$; $y_1 \in \{y_2\}.\varphi_\eta$ implies $\{y_1\}.\eta \subseteq \{y_2\}.\eta$; so $\{y_0\}.\eta \subseteq \{y_2\}.\eta$ and thus $y_0 \in \{y_2\}.\varphi_\eta$. But this contradicts the assumption that $y_0, y_2 \in I$ which is irreducible. Thus $i(y_0) = y_2$ is part of the desired isometry, i. □

In Figure 1, the graph G of 18 vertices is reduced to $I = G.\omega$ with 10 remaining vertices. In G, the dashed lines denote the vertices that belong to $2', 3', 15'$ and

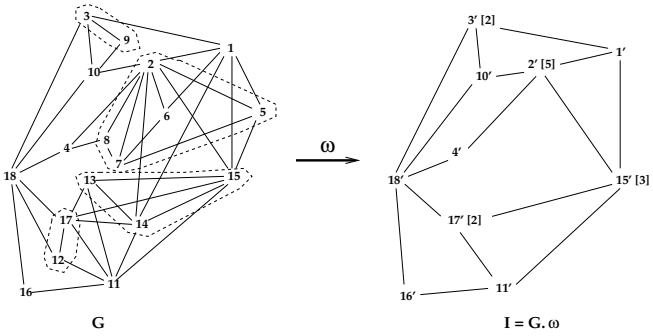

Figure 1: Reduction, ω, of a graph G to its spine I.

$17'$ respectively. For those vertices, $|\{y\}.\beta|$ is represented by $[n]$ in the figure. For all other vertices, $|\{x\}.\beta| = 1$. (These values will be used in Section 2.2.)

[6]This procedure has been quite effective reducing large graphs $|V| \geq 1,000$, with at worst 6 iterative sweeps of V.

Reduction, $G.\omega$, of G to an irreducible graph I has a number of interesting properties. Let $\sigma(x, z)$ denote a **shortest path** between x and y, that is $< x, \ldots, y >$ is of minimal length. $\sigma(x, y)$ need not be unique. By the **distance between** x and y, $d(x, y)$, we mean the length of any shortest path $\sigma(x, y)$.

Proposition 2.3. *Let $\sigma(x, y)$ denote a shortest path between x and y. If z belongs to y then $z \notin \sigma(x, y)$.*

Proof: Suppose not, that $\sigma(x, y) = < x, \ldots, x_n, z, y >$. Since z belongs to y, $z \in \{y\}.\varphi_\eta$, or $\{z\}.\eta \subseteq \{y\}.\eta$ implying $x_n \in \{y\}.\eta$ yielding a shorter path $< x, \ldots, x_n, y >$. □

In other words, z can be removed from V with the certainty that if there was a path from some vertex x to y through z, it was not a shortest path. Such vertices can be iteratively removed from V without changing connectivity, or distances, between retained vertices. In Figure 2, $d(x, y) = 2$. Any two of $\{z_0, z_1, z_2\}$ may be thought to belong to the remaining vertex yielding 3 possible isomorphic

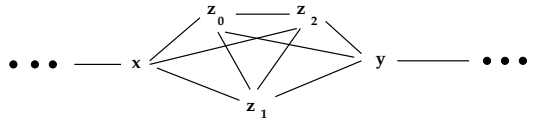

Figure 2: Shortest paths $\sigma(x, y)$ between x and $y \in G$.

spines, I_0, I_1, I_2, while retaining a shortest path. Figure 2 also illustrates a configuration that can occur in the proof of Proposition 2.2.

A central quest in the analysis of social networks is the identification of its "important" vertices. In social networks, "importance" is often defined in terms of the "hubs" of a network (a concept we explore more fully in Section 3, as well as with respect to the path structure of the network [9].

Those vertices $C_D = \{y \in V\}$ for which $\delta(y) = \sum_{s \neq y} d(s, y)$ is $minimal$ have traditionally been called the **center** of G; they are "closest" to all other vertices. It is well known that this subset of vertices must be edge connected [11]. One may assume that these vertices, in the "center", of a graph are "important" vertices.

Alternatively, one may consider those vertices which "connect" many other vertices, or clusters of vertices, to be the "important" ones. Let $\sigma_{st}(y)$ denote the number of shortest paths $\sigma(s, t)$ containing y; then those vertices y for which $\sigma_{st}(y)$ is $maximal$ are those vertices that are involved in the most connections.

Let $C_B = \{y \in V\}$, for which $\sigma_{st}(y)$ is maximal. This is called **betweenness centrality** [6, 9].

It can be shown that vertices with minimal distance and maximal betweenness measures will *almost always* be found in, or adjacent to, the irreducible spine I. This is non-trivial because it need not always be true. One problem is that, we may have several isomorphic spines, $I_1, \ldots, I_k$, so we can only assert that $C_D \cap I_j$ and $C_B \cap I_j$ are non-empty for some $1 \leq j \leq k$. Second, there exist pathological cases where the centers are completely disjoint from I. However,

Proposition 2.4. *Let I be an irreducible spine of a graph G with centers C_D and C_B. If for all $y \in I$, $\sum_{x \in I, x \neq y} |\{x\}.\beta| \geq |\{y\}.\beta|$ and $\sum_{x \in I, x \in \{y\}.\eta} |\{x\}.\beta| \geq |\{y\}.\beta|$ then there exist $x_i \in I$ and $y_j \in I$ such that $x_i.\eta \cap C_D$ and $y_j.\eta \cap C_B \neq \emptyset$.*

Proof: A tedious proof can be found in [23] □

The conditions of Proposition 2.4 are sufficient; but are by no means necessary. In practice, one really only needs that I be sufficiently large, and that its reduced subsets not be too unbalanced.

It turns out that the irreducible spine, $I = G.\omega$, is a rather good abstraction of a symmetric graph, G.

2.2. Graph Expansion, ε

It is fairly easy to define the treatment of edges in an operation, such as ω, that contracts a graph. If $Y \xrightarrow{f} \emptyset$, then all edges $\{y, z\}$ such that $y \in Y, z \in Y.\eta$ can be deleted. Expanding a graph, $\emptyset \xrightarrow{g} Y'$, presents more problems. How is Y' to be embedded in G'? One option is to employ an *expansion grammar* ε, such as explored in [22]. Expansion grammars are quite different from phrase-structured grammars in which a non-terminal symbol A is expanded with a rewrite rule of the form $A \rightarrow \sigma$, *c.f.* [26]. The problematic aspect of a phase-structured grammar, explored by Ehrig in [8], is how will the right side σ of the rewrite rule be embedded in the growing, non-linear structure.

In an expansion grammar, a subset Y of a growing structure is first identified to be the neighborhood of the new element z'. That is $\{z'\}.\eta' = Y \subseteq V$ in the rewritten structure. More precisely, $\varepsilon_i : (V_i, E_i) \rightarrow (V_{i+1}, E_{i+1})$ where $V_{i+1} = V_i \cup \{z_i'\}$, $E_{i+1} = E_i \cup \{\{y_k, z_i'\}, y_k \in Y \subseteq V_i\}$ and $\varepsilon_i : \emptyset = \{z_i'\}$.

For our purposes, we use a simple rewrite rule, r, in which Y can be any subset of an existing neighborhood $\{y\}.\eta$, $y \in V_i$. Such a rewrite operation is implemented by the following code.

```
for all y in V
    {
    while (|{y}.beta| > 1)
        {
        create new node z;
        S = choose_random_in ({y}.nbhd);
        {z}.nbhd = S;
        k = random_int(1, |{y}.beta|-1);
        |{y}.beta| = |{y}.beta| - k;
        |{z}.beta| = k;
        add {z} to N;
        }
    }
```

Figure 3 shows one possible application of this expansion grammar ε to the graph I of Figure 1. Here, the rewrite rule r has been used 8 times, to create

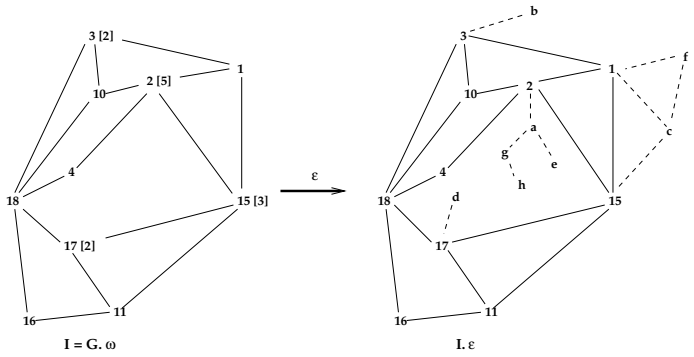

Figure 3: A member of $I.\varepsilon$ where $I = G.\omega$ in Figure 1.

$a, b, \ldots h$. The vertex d is generated by r using the neighborhood $Y = \{17\} \subseteq \{15, 17, 18\} = \{17\}.\eta$. So, $\{d\}.\eta' = \{17\} \subset \{17\}.\eta$. The new vertex c was attached to $Y = \{1, 15\} \subset \{1, 2, 3, 15\} = \{1\}.\eta$; and f was later attached to $\{1, c\} \subset \{1\}.\eta$. Beause $|\{2\}.\beta| = 5$, five vertices $\{2, a, e, g, h\}$ now belong to $\{2\}$ in $I.\varepsilon$.

2.3. The Inverse Set, ω^{-1}

The two procedures ω and ε are intertwined. The requirement that $\{z'\}.\eta = Y \subseteq \{y\}.\eta$ ensures that if ω is applied to $I.\varepsilon$, z' will at some iteration belong to y. Thus, if I is irreducible, $I.\varepsilon.\omega = I$ This characteristic is evident in Figure 3 where

b will be reduced by 3, $etc.$ It is also true for the graph $I.\varepsilon$ of Figure 4 as well. Consequently, ω is a right-inverse of ε over the subspace of irreducible undirected graphs. The inverse of ω, that is $G.\omega.\omega^{-1}$ is the collection of all undirected graphs $\{G_k\}$ such that $G_k.\omega = I = G.\omega$. Each invocation of the non-deterministic procedure ε is single-valued; but ε is not a well defined function. The execution of ε will yield just one graph, $G_k \in G.\omega.\omega^{-1}$.

In the expansion operation ε, the choice of $y \in V_i$ and the choice of $Y \subseteq \{y_i\}.\eta$ are completely arbitrary. Given different choices for y and Y yields Figure 4 which seems to be a far more interesting graph. Both Figures 3 and 4 were

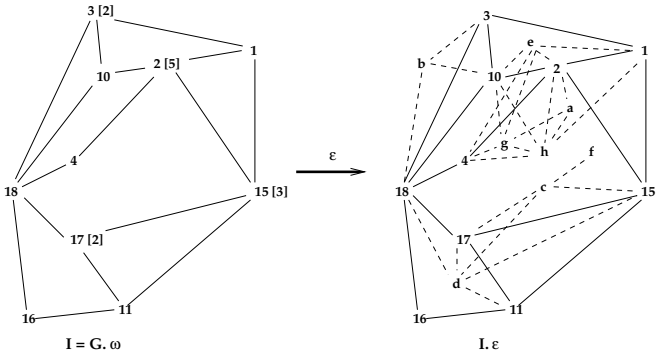

Figure 4: Another graph $I.\varepsilon$ in $I.\omega^{-1}$.

generated by a computer version of ε using a random number generator.

3. How Similar are Similar Graphs

In social network analysis there is considerable emphasis on the *hubs* of the network [13]. These vertices, which characteristically have high degree, are regarded as *authorities* that are frequently referenced. (In network analysis, it is customary to call the vertices of the graph structure its "nodes".) Consequently, one of the more common similarity measures is a comparison of node degree, that is the number of related nodes or $|\{y\}.\eta| - 1$. Do the hubs of one network correspond to the hubs in the other network? All of the similarity measures in [12, 19, 30] consider node degree.

The vertex of each network graph are sorted in descending order of their degrees. Those vertices common to both networks are then rank ordered with respect to their degree. Network hubs have the highest (most significant) rank. The Spear-

man coefficient, ρ, can then be calculated as usual by:

$$\rho(G_1, G_2) = 1 - \frac{6 \cdot \sum_{all\ vertices\ y}(rank_1(y) - rank_2(y))^2}{n(n^2 - 1)} \tag{4}$$

where n denotes the number of vertices common to both graphs and the rank of vertices with the same degree is the average of all ranks of vertices with that degree.

Table 1 illustrates typical Spearman coefficients, ρ, for various initial network graphs, G, and some expanded version, that is $G.\omega.\varepsilon$. The columns labeled n_G and

Graph	n_G	e_G	hubs	n_ω	e_ω	n_ε	e_ε	ρ
Figure 3	18	41	6	10	16	18	30	0.7614
Figure 4	18	41	6	10	16	18	46	0.9324
random 1	200	310	15	119	223	200	327	0.9353
random 2	200	355	14	147	302	200	350	0.9648
random 3	200	396	15	161	356	200	443	0.9687

Table 1: Similarity of original vs. reconstructed graphs based on vertex degrees.

e_G denote the number of nodes and edges in the original graph G. The columns labeled n_ω, e_ω, n_ε and e_ε denote the number of nodes and edges in $G.\omega$ and $G.\omega.\varepsilon$ respectively. The final column is the Spearman coeficient, ρ between G and $G.\omega.\varepsilon$ based on the rank correlation between the individual vertex degrees. The first row describes properties of the small undirected graph, G and its spine $I = G.\omega$, shown in Figure 1, followed by it's reconstruction $G.\omega.\varepsilon$ shown in Figure 3. We consider it to have 6 hubs, that is vertices with $|\{y\}.\eta| \geq 6$. We think the low similarity measure is due to the relatively few reconstructed edges and to the inherent variability of statistical results with small graphs. The second row compares yet another reconstruction of the graph G in Figure 4. It seems much more typical. By recording $\{y\}.\beta$, our reconstruction code knows how many nodes were removed in the reduction process, and roughly where. Consequently, n_ε always equals n_G and so the reconstructed networks can be rather close.

To further illustrate the similiarity of reconstructed (expanded) graphs in ω^{-1} with their source, we generated random 200 vertex networks with roughly 300, 350, and 400 links, or edges, in each. The random process was skewed so that 15 vertices received probabilistic preference, and thus would emerge as hubs. These constitute the final three rows of Table 1.

4. Cycle Systems

The irreducible spine in Figures 1, 3 and 4 consisting of 7 cycles of length $< 3, 3, 3, 4, 4, 4, 5 >$ each. All irreducible spines must be cycle systems such as this.

Proposition 4.1. *If $y \in I = G.\omega$ then y is an element of a chordless cycle of length ≥ 4.*

Proof: Let $y_0 = y \in I$ and let $y_1 \in \{y_0\}.\eta$. Since I is irreducible, $\{y_1\}.\eta \not\subseteq \{y_0\}.\eta$, so $\exists y_2 \in \{y_1\}.\eta$, $y_2 \notin \{y_0\}.\eta$. Again, since I is irreducible, $\exists y_3 \in \{y_2\}.\eta$, $y_3 \notin \{y_1\}.\eta$. else $\{y_2\}.\eta \subseteq \{y_1\}.\eta$ contradicting irreducibility. But, possibly, $y_3 \in \{y_0\}.\eta$ in which case $< y_0, y_1, y_2, y_3, y_0 >$ is the desired chordless cycle.
If $y_3 \notin \{y_0\}.\eta$, irreducibility requires that $\exists y_4 \in \{y_3\}.\eta$, $y_4 \notin \{y_2\}.\eta$. Again possibly $y_4 \in \{y_1\}.\eta$ or $y_4 \in \{y_0\}.\eta$, yielding a chordless cycle of length 4, or of length 5.
Otherwise $\exists y_5 \in \{y_4\}.\eta$, $y_5 \notin \{y_3\}.\eta$, and so on. Since V is finite, this construction must terminate with the desired chordless cycle. □

So, essentially, irreducible spines, I, must be cyclic structures. Even though there are 3 triangles in Figures 1, 3 and 4 one can verify that the proposition is still true.

For any cycle, C_i in a graph $G = (V, E)$, let $\dot{C}_i = \{y_j\}, y_j \in V$ and let $\bar{C}_i = \{(y_j, y_{j+1}\}, (y_j, y_{j+1}) \in E$. That is, $\dot{c}_i$ denotes the edges in the cycle and $\bar{C}_i$ denotes its edges. It is often clearer to describe cycle by enumerating their constitutent nodes; but unless the cycle are "chordless" this need not be a unique definition.

We define cycle composition $C_i \circ C_k$ to be

$$C_i \circ C_k = (\bar{C}_i \cup \bar{C}_k) \sim (\bar{C}_i \cap \bar{C}_k) \tag{5}$$

For example, in the running cycle system $I = G.\omega$ of figures 1, 3, 4, 5 and 6, let $C_i =< 2, 4, 18, 10 >$ and $C_k =< 2, 15, 17, 18, 4 >$ then readily $C_i \circ C_k =< 2, 15, 17, 18, 10 >$

The first four propositions are easily derivable from the definition (5).

Proposition 4.2. *For all i, k, $C_i \circ C_k = C_k \circ C_i$.*

Proposition 4.3. *For all i, k, m, $C_i \circ (C_k \circ C_m) = (C_i \circ C_k) \circ C_m$.*

Proposition 4.4. *For all i, $C_i \circ C_i = C_\emptyset$ where $\dot{C}_\emptyset = \emptyset$ and $\bar{C}_\emptyset = \emptyset$.*

Proposition 4.5. *For all i, $C_i \circ C_\emptyset = C_i$.*

Proposition 4.6. *] If $C_m = C_i \circ C_k$ then $C_k = C_i \circ C_m$.*

Proof: Let $C_m = C_i \circ C_k$, then $C_k = C_\emptyset \circ C_k = (C_i \circ C_i) \circ C_k = C_i \circ (C_i \circ C_k) = C_i \circ C_m$. □

Proposition 4.7. *If $C_i \neq C_k$ then $C_i \circ C_k \neq C_\emptyset$ for $i \neq k$.*

Proof: Suppose $C_i \circ C_k = C_\emptyset$, where $i \neq k$, then $C_k = C_\emptyset \circ C_k = (C_i \circ C_i) \circ C_k = C_i \circ (C_i \circ C_k) = C_i \circ C_\emptyset = C_i$, or $C_i = C_k$ contradicting the condition. □

4.1. Cycle Systems as Matroids

Let $\mathcal{C}$ denote the collection of all cycles in a graph. A set $\mathcal{S} = \{C_i\} \subseteq \mathcal{C}$ of non-empty cycles is said to be **dependent** if there exists $C_m \in \mathcal{S}$ such that $C_m = C_i \circ \ldots \circ C_k$, where $C_i \ldots, C_k \in \mathcal{S}$. If $\mathcal{S}$ is not dependent, it is said to be **independent**. Any cycle can be a member of an independent set.

Let $Y = \{C_1, \ldots, C_k\}$ be a set of cycles. By the **span** of Y, denoted $Y.\sigma$, we mean the set of all cycles $\{C_m\}$ such that $C_m = C_i \circ \ldots \circ C_j \circ \ldots \circ C_k$, where $C_j \in Y$.[7] It is convenient to introduce a bit of new notation. Let $Y \subseteq \mathcal{C}$ be some collection of cycles. We let $Y\circ$ denote the composition of all the cycles in Y. Using this notation, we define $Y.\sigma$ to be $Y \cup \{X\circ, X \subseteq Y, |X| > 1\}$ Readily, $Y \subseteq Y.\sigma$.

Proposition 4.8. *Let $Y \subseteq \mathcal{C}$ be an independent collection of cycles. If $C_i = X\circ$, $X \subseteq Y$, then X is unique (in Y).*

Proof: Let $C_i = Y\circ$. Suppose there exists $Z \subseteq Y, Z \neq X$ such that $C_i = Z\circ$. $|Z| \geq 2$ else $Z = C_i$ contradicting assumed independence of Y. Let $C_k \in Z$, then $C_i = C_k \circ (Z{\sim}C_k)\circ$. By Prop. 4.6, $C_k = C_i \circ (Z{\sim}C_k)\circ$, and $C_k = (X\circ) \circ (Z{\sim}C_k)\circ$ again contradicting the independence of Y. □

[7] In graph theory, the term "span" usually refers to a tree whose nodes include all $y \in N$. Since a tree has no cycles, it has no connection to our usage which is taken from the notion of spanning vector spaces.

If an independent collection Y of cycles also spans $\mathcal{C}$, we say Y is a **basis** for $\mathcal{C}$ and normally denote such a set by $\mathcal{B}$. There can be many bases $\mathcal{B}_1, \mathcal{B}_2 \ldots$ for $\mathcal{C}$

Let $\mathcal{B}$, where $|\mathcal{B}| = n$, be a basis for $\mathcal{C}$. So $\mathcal{B}$ spans $\mathcal{C}$. By Prop. 4.8, every subset $Y \subseteq \mathcal{B}$ is unique, so the cardinality of the cycle system, $|\mathcal{C}| = |\{Y \subseteq \mathcal{B}\} = \sum_{1 \leq k \leq n} C(n,k) = 2^n$. In effect, Prop. 4.8 assures that the cardinality of every basis must be the same. That all bases have the same cardinality is a principal result in matroid theory. Here it follows from Proposition 4.8 before we establish that a cycle system $\mathcal{C}$ actually is a matroid.[8] We show that now.

Proposition 4.9. *The spanning operator, σ is a closure operator over sets Y of cycles.*

Proof: Readily, σ is expansive and monotone.
Let Y be a set of cycles $\{C_i\}$. Suppose $C_m \in Y.\sigma.\sigma$ implying that there exists some sequence $1 \leq i \leq k$ such that $C_m = C_1 \circ \ldots \circ C_i \circ \ldots \circ C_k$, where $C_i \in Y.\sigma$, $1 \leq i \leq k$. Hence $C_i = C_{i_1} \circ \ldots \circ C_{i_n}$ where $C_{i_j} \in Y$.
Thus, substituting for each i in the sequence above for C_m, we get $C_m = (C_{1_1} \circ \ldots \circ C_{1_n})$ $\circ\, (C_{2_1} \circ \ldots \circ C_{2_n}) \circ \ldots \circ (C_{k_1} \circ \ldots \circ C_{k_n})$ implying $C_m \in Y.\sigma$. □

A closure system is said to be a **matroid** if it satisfies the Steinitz-MacLane **exchange** axiom [10, 14, 29], that is:

if $x, y \notin Y.\varphi$ and $y \in (Y \cup x).\varphi$ then $x \in (Y \cup y).\varphi$.

If φ satisfies the *anti-exchange* axiom [21], that is:

if $x, y \notin Y.\varphi$ and $y \in (Y \cup x).\varphi$ then $x \notin (Y \cup y).\varphi$

then the system is called an *antimatroid*.

Proposition 4.10. *Let $\mathcal{C}$ be a normal cycle system and let σ be the spanning operator. The system $(\mathcal{C}, \sigma)$ satisfies the Steinitz-Maclane exchange axioms and is thus a matroid.*

Proof: By Prop. 4.9, σ is a closure operator.
Let $C_i, C_k \not\subseteq Y.\sigma$ where $Y = \{\ldots, C_j, \ldots\}$. Suppose $C_k \in (Y \cup C_i).\sigma$ implying that $C_k = C_i \circ (\ldots C_j \ldots) = C_i \circ C_m$ where $C_m \in Y.\sigma$. Consequently, by Prop. 4.6, $C_i = C_k \circ C_m$ or $C_i \in (Y \cup C_k).\varphi$. □

[8] Welsh [29] defines the "cycle matroid" of a graph in which any cycle is a dependent set of edges, while trees are independent. Any maximal subtree is a basis. Our "cycle matroid" is quite different.

Since $\mathcal{C}$ is a matroid, the cardinality of every maximal independent set is fixed and this number, r, is the **rank** of the system. This was established earlier by Proposition 4.8. Since any cycle system $\mathcal{C}$ constitutes a matroid, it satisfies the following fundamental basis exchange theorem.

Proposition 4.11. *Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be any two bases of $\mathcal{C}$, and let $C_i \in \mathcal{B}_1$. Then there exists $C_k \in \mathcal{B}_2$ such that $(\mathcal{B}_1 \sim C_i) \cup C_k$ is a basis of $\mathcal{C}$.*

Matroids and vector spaces are more often characterized by this ability to arbitrarily exchange basis elements. Proposition 4.6 effectively provides this ability.

Cycle systems derived from arbitrary undireced graphs by the reduction operator ω are most interesting mathematical objects. But, we will not develop these matroid properties further in this paper, except use the concept of rank, r, in the next section.

4.2. Normal and Cubic Cycle Systems

Every cycle system can be transformed into a normal form in which every node is either binary or triadic (has degree 2 or 3). See Figure 5 for the normal form of Figures 1, 3, 4. It is now easy to see that every cycle system in normal form is

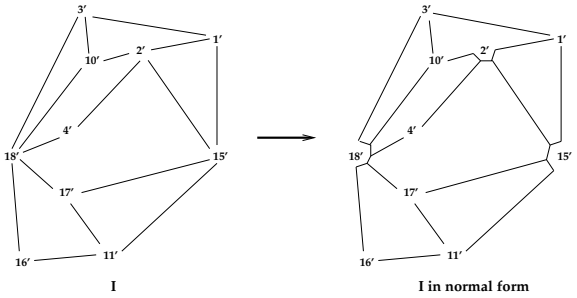

Figure 5: The normal form of Figure 1(b) and 3(a).

homeomorphic (topologically equivalent) to a cubic graph, where all nodes have degree 3, by just eliminating the binary nodes. We should note that some cycle systems will give rise to multi-edges in the cubic graph. Figure 6 is the cubic graph corresponding to Figure 5.

There is a considerable literature devoted to cubic graphs, of which [2, 3, 15, 16] is just a sample. Of most relevance here is an enumeration of cubic graphs in Table II. based on rank, and found in [7]. The details in their description of the generation process provides an excellent primer on the anatomy of cubic graphs.

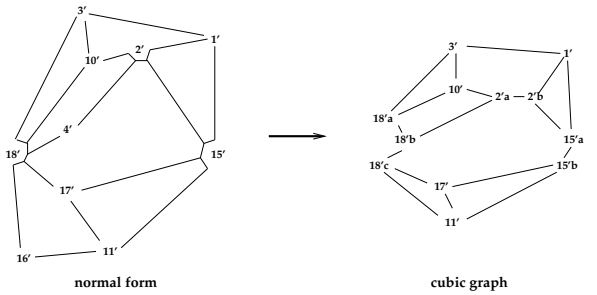

Figure 6: The cubic graph equivalent to I in Figure 1(b).

r	n	connected cubic graphs
3	4	1
4	6	2
5	8	5
6	10	19
7	12	85
8	14	509
9	16	4,060
10	18	41,301
11	20	510,489
12	22	7,319,447
13	24	117,940,535
14	26	2,094,480,864
15	28	40,497,138,011
16	30	845,480,228,069
17	32	18,941,522,184,590

Table 2: Number of connected cubic graphs of rank r.

The irreducible spine I in Figures 1, 3, and 4 has rank 7. So Figure 1(b) is representative of one of 85 distinct equivalence classes having irreducible spines with 7 independent cycles.

We say a cubic graph, such as Figure 6 defines a **genus** consisting of those graphs whose irreducible spines have the same cycle structure. Readily, genus also defines an equivalence relation on the space of all undirected graphs. It is a coarser equivalence than that created by irreducible spines, I, because if G_1 and G_2 have the same spine I, they must be topologically equivalent the same cubic graph, and thus belong to the same genus of graphs.

The concept of "genus" already exists in graph theory. It denotes the minimal number of "handles" on an orientable sphere needed to embed the graph without edge crossing [4, 5]. Graphs of genus 0 are planar. Readily, this notion of a numerical genus is an equivalence as well; one that is still coarser than our cubic graph genus. However, calculating the numerical genus of a graph involves determining an orientable embedding which is computationally most difficult, even for relatively small graphs. Both the calculations for irreducible similarity and cubic graph similarity are local, fast, and easily parallalizable. They are practical.

5. Discussion

Crude measures of graph similarity can be valuable. But, unless one is very familiar with the statistical process generating the measure, one doesn't really know what makes the two graphs similar. Is it that they have similar profiles of vertex degree distribution? or of average path length distributions? Any measure of similarity based on path structure, or path length, will be computationally expensive.

The concept of irreducible similarity developed in this paper tells you a great deal about "similar" graphs. They have essentially the same path structure, the same "centers" which must be located on the common irreducible spine (or one link away), and vertex degree distributions that are very close. Similarity within a cubic graph genus encompasses a much broader collection of "similar" graphs, but tells much less about them. They have the same rank (maximal number of independent cycles) and they are connected in the same fashion.

But, both similarities are equivalence relations. Two graphs G_1, G_2 are either similar, or not.

If not, no information whatever is conveyed regarding either their similarity or their differences. It is possible that by reducing both graphs and comparing their spines, $I_1 = G_1.\omega$ and $I_2 = G_2.\omega$, one could broadly estimate the similarity between members of the two distinct equivalence classes in the unbounded collection of all undirected graphs. We have not yet done this.

References

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

[2] R.E.L. Alfred, S. Bau, D.A. Holton, and Brendan D. Mckay. NonHamiltonian 3-connected cubic planar graphs. *SIAM J. Discrete Math.*, 13(1):25–32, 2000.

[3] Pouya Baniasadi, Vladimir Ejov, Jerzy A. Filar, and Michael Haythorp. Genetic theory for cubic graphs. *arXiv*, 1, May 2018. arXiv:1209.5514v1.

[4] Mehdi Behzad and Gary Chartrand. *Introduction to the Theory of Graphs*. Allyn and Bacon, Boston, 1971.

[5] Béla Bollobás. *Modern Graph Theory*. Springer, New York, 1998.

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

[7] Gunnar Brinkmann, Jan Goedgebeur, and Brendan D. Mckay. Generation of Cubic Graphs. *Discrete Math. and Theoretical Comp. Sci., DMTCS*, 13(2):69–79, 2011.

[8] H. Ehrig, M. Pfender, and H. J. Schneider. Graph Grammars: An Algebraic Approach. *IEEE Conf. SWAT*, 1973.

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

[10] George Gratzer. *General Lattice Theory*. Academic Press, 1978.

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

[12] G. Jeh and J. Widom. SimRank: A measure of structural context similarity. In *Proc. 8th Intern. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 538–543, Edmonton, 2002.

[13] Jon M. Kleinberg. Hubs, authorities and communities. *ACM Computing Surveys*, 31(4), Dec. 1999. Doi.acm.org/10.1145/345966.345982.

[14] Bernhard Korte and László Lovász. Intersections of Matroids and Antimatroids. *Discrete Math.*, 73:143–157, 1988/89.

[15] Charles H. C. Little. A characterization of planar cubic graphs. *J. Comb. Theory, Series B*, 29(2):185–194, Oct. 1980.

[16] Brendan D. Mckay and Nick. C. Wormald. Automorphisms of random graphs with specified degrees. *Ars Combinatoria*, 4:325–338, 1984.

[17] Oystein Ore. Mappings of Closure Relations. *Annals of Math.*, 47(1):56–72, Jan. 1946.

[18] Ratko Orlandic, John L. Pfaltz, and Christopher Takylor. 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.

[19] Panagiotis Papadimitriou, Ali Dasdan, and Hector Garcia-Molina. Web graph similarity for anomaly detection. *J. Internet Serv. Appl.*, pages 19–30, 2010. Doi: 10.1007/s13174-010-0003-x.

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

[21] John L. Pfaltz. Closure Lattices. *Discrete Mathematics*, 154:217–236, 1996.

[22] John L. Pfaltz. Neighborhood Expansion Grammars. In *Intern'l Workshop on Theory and Application of Graph Transformation, TAGT'98*, Lecture Notes in Computer Science, #1764, pages 30–44, Paderborn, Germany, Nov. 1998. Springer-Verlag.

[23] John L. Pfaltz. The Irreducible Spine(s) of Discrete Networks. In Xuemin Li, Yannis Manolopoulos, Divesh Srivastava, and Guangyan Huang, editors, *Web Information Systems Engineering - WISE 2013*, volume LNCS # 6984, Part 2, pages 104–117), Nanjing, PRC, Oct. 2013.

[24] John L. Pfaltz. The Role of Continuous Processes in Cognitive Development. *Mathematics for Applications*, 4(4):129–152, 2015. doi:10.13164/ma.2015.11.

[25] John L. Pfaltz. Using Closed Sets to Model Cognitive Behavior. In Tapabrata Ray, Ruhul Sarker, and Xiaodong Li, editors, *Proc. Australian Conf. on Artificial Life and Computational Intelligence (ACALCI 2016)*, volume LNCS 9592, pages 13–26, Canberra, ACT, 2016.

[26] Gregorz Rozenberg, editor. *The Handbook of Graph Grammars*. World Scientific, 1997.

[27] Harold S. Stone. *Discrete Mathematical Structures and Their Applications*. SRA, Palo Alto, 1973.

[28] Josef Šlapal. A Galois Correspondence for Digital Topology . In K. Denecke, M. Erné, and S. L. Wismath, editors, *Galois Connections and Applications*, pages 413–424. Kluwer Academic, Dordrecht, 2004.

[29] D.J.A. Welsh. *Matroid Theory*. Academic Press, 1976.

[30] Laura A. Zager and George C. Verghese. Graph similarity scoring and matching. *App. Math. Letters*, 21:86–94, 2008. Doi: 10.1016/j.aml.2007.01.006.

[31] Yuanyuan Zhu, Lu Qin, Jeffrey Xu Yu, and Hong Cheng. Finding Top-K Similar Graphs in Graph Databases. *EDBT 2011*, March 2012.