Report on “Orientability of Phylogenetic Network Graphs” by Ethan Cecchetti.

submitted to *Rose-Hulman Undergraduate Mathematics Journal*.

November 6, 2007

*Summary*:

The author characterizes *trivalent graphs* (connected graphs whose vertices have degree 1 or 3) that can be oriented into *network graphs* (directed graphs such that no vertex has all edges directed inward or all edges directed outward and such that no directed circuit exists). Network graphs have application in phylogenetic reconstruction when modeling species evolution where both speciation and hybridization events can occur. The author's characterization is elegantly stated, as is the proof.

*Recommendation*:

The results of this paper are correct and interesting, and in general well-presented. There are several places where definitions are unclear, making the rest of the paper confusing. However, I think all of these places can be easily fixed. Therefore, subject to the suggestions below, I recommend acceptance of this paper for publication in the *Rose-Hulman Undergraduate Mathematics Journal*.

*Specific Comments*:

*Abstract*:

p 1, line 1: “the genetic history of species…”.

p 1, line 4: Remove the comma after trivalent.

*Section 1. Introduction*:

p 1, line 1: Usually, the vertices represent species, and edges indicate that species $x$ is an immediate ancestor of species $y$.

p 1, lines 10–13: In the three sentences starting with “Phylogenetic networks…”, it is unclear when you are defining *network graphs*, and when you are defining *phylogenetic networks*. I recommend combining the two sentences on lines 12–13 as follows: “In a network graph, circuits are allowed, edges are given directions, and every internal vertex is trivalent (has three edges connected to it).”

p 1, line 16: “…but later hybridizes”: this clause is confusing; what is its subject? Is the species hybridizing with itself? Are its two direct descendants hybridizing, implying that the circuit is of length 4?

p 2, line 9: Acknowledgements are usually placed in a separate section at the end of the paper before the references.

*Section 2. Terminology*:

p 2, line 3: Defining edges as “lines” implies that a graph is always drawn, but this need not be the case. It is better to define edges as unordered pairs of vertices and then comment that we often think of graphs as drawn points and lines.

p 2, line 4: Remove extra “and”.

p 2, line 4: Defining a circuit as a “loop” is confusing, since the term *loop* usually refers to a circuit consisting of one edge. This common definition of loop is used in the last proof.

p 2, line 5: “…to length one…” should be “…of length one…”.

p 2, line 13: Change “all nodes” to “all internal nodes” for emphasis.

p 2, line 13: All graphs in the paper are implicitly assumed to be connected; otherwise, some of the results are false. It would be good to explicitly state that all graphs are assumed to be connected. One convenient place to do that would be in the definition of trivalent graph. However, it would be useful to remind the reader that the graphs under consideration are connected at key points throughout the paper, such as Case 1 of the proof of Theorem 4.3.

p 2, line 15: Remove comma after "orientation".

*Section 3. Non-Orientable Graphs:*

p 3, Thm 3.1: Remove commas around "$\Gamma$"; they are not necessary. Remove the commas in other instances as well.

p 4, Thm 3.3: Before the theorem, define "inward" and "outward"; it is not clear if an inward directing edge is oriented towards the leaf or away from it.

p 5, lines 2–3: Remove commas around "$\Gamma$" and "$\Gamma'$".

p 5, lines 2–3: The definition of $\Gamma'$ is fundamental to the proofs that follow. However, it is unclear what happens to the other edges connected to $E$'s endpoints when those vertices are removed. Figure 6 makes it appear that the endpoints of $E$ have two common neighbors, and that $E$ must be the diagonal of this "diamond." I assume that is not what is meant, but a much clearer definition is needed.

p 5, Lemma 4.1: Give the relationship between $\Gamma$ and $\Gamma'$ in the statement of the lemma: "Suppose that $\Gamma'$ is formed from $\Gamma$ by .... If $\Gamma'$ can be oriented as a network graph, ...".

p 5, proof of Lemma 4.1, line 7: The claim is made that there is a directed path from $V$ to $W$ in $\Gamma'$, but $V$ and $W$ are not vertices in $\Gamma'$.

p 5, proof of Lemma 4.1, line 8: "...then $\Gamma'$ already has a directed circuit."

p 6, Figure 7: Labels "$\Gamma_1$" and "$\Gamma_2$" on the appropriate graphs would be helpful.

p 7, lines 1–2: The definition of "illegal subgraph" is confusing. Saying that a subgraph $\Psi$ is trivalent means that $\Psi$ is a trivalent graph. If $\Psi$ has no leaves, then every vertex in $\Psi$ has degree 3 in $\Psi$, and hence there can be no edge from vertices of $\Psi$ to the rest of the graph $\Gamma$. I think the definition of illegal subgraph should be: "An *illegal subgraph* is a trivalent subgraph with exactly one leaf." The leaf edge is then the separating edge discussed in the next line. This definition also makes it clear why illegal subgraphs are a problem, since then Corollary 3.4 immediately applies.

p 7, proof of Thm 4.3, Case 1, line 4: "...a unique directed path from $V$ to..."

p 7, proof of Thm 4.3, Case 2, line 4: Here "loop" is used to mean a circuit of length one. It might be clearer to remind the reader at this point that loops are not allowed in network graphs, since their orientation creates a directed circuit.

p 7, proof of Thm 4.3, Case 3, line 2: Remove comma before "$E_1$".

p 8, line 2: Again, "loop" is used to mean a circuit of length one.

p 8, line 5: "...would separate the graph."

p 8, line 6: "one of the resulting smaller graphs, $\Gamma_{ill}$, has no leaves, while the other, $\Gamma_2$, has all the leaves from $\Gamma$."

p 8, line 9: "Because $\Gamma$ has no..."