Cumhuriyet Science Journal

Cumhuriyet Sci. J., 44(1) (2023) 130-142
DOI: https://doi.org/10.17776/csj.1071349

| csj.cumhuriyet.edu.tr | Founded: 2002 ISSN: 2587-2680 e-ISSN: 2587-246X Publisher: Sivas Cumhuriyet University

The Structure of Level-2 Semi-directed Binary Phylogenetic Networks

Nihan Özbaltan [1,a,*]

[1] *Department of Computer Engineering, Faculty of Engineering and Architecture, Bakırçay University, İzmir, Türkiye.*
**Corresponding author*

Research Article

History
Received: 10/02/2022
Accepted: 13/11/2022

ABSTRACT

Phylogenetic trees describe evolution but due to hybridization events, recombination events or lateral gene transfer, it can be represented as a phylogenetic network. In phylogenetic networks, some of the branches of tree combine and create a reticulation node. Level of a network is decided to look at how many nodes in a connected component in a network. In this research, Int this paper, the structure of directed and undirected level-2 networks and how the can be decomposed into level-2 generators is studied.

Keywords: Mathematical biology, Phylogenetics, Phylogenetic networks, Biostatistics.

[a] *nihan.ozbaltan@bakircay.edu.tr* *https://orcid.org/0000-0003-0191-312X*

Introduction

Phylogenetics is an interdisciplinary study, which is the interface of biology, mathematics, and computer science. There are different methods to construct phylogenetic trees: Maximum Parsimony (MP), Maximum Likelihood (ML), Bayesian methods, Distance-based methods and quartet-based methods. However, phylogenetic trees are not enough to explain evolution due to hybridization events, recombination events or lateral gene transfer. Phylogenetic networks are used to generalize the tree model of evolution. In a phylogenetic nework, addition branches join vertices, which are already connected by a path, defined as a reticulation. Quartet methods apply to the construction of unrooted evolutionary trees on four leaves.

In this study, phylogenetic network is a directed acyclic graph containing a single root vertex (indegree 0 and outdegree 2), leaf vertices (indegree 1 and outdegree 0) and reticulation vertex (indegree 2 and outdegree 1). A graph is biconnected, which does not contain any vertex whose removal disconnects the graph. Level of a phylogenetic network is the greatest number of reticulation vertices in a biconnected component of a graph. Phylogenetic tree is a level-0 network. My focus on this paper is level-2 network, which is illustrated in Figure 1. Each grey blob shows the biconnected component. Gambette et al. shows the structure of level-k phylogenetic networks and showed that how they can be decomposed into level-k generators [1]. Jansson and Sung constructed level-1 networks with a given set of triplets [2]. Van Iersel et al. extended Jansson's work to construct level-2 phylogenetic networks from triplets [3]

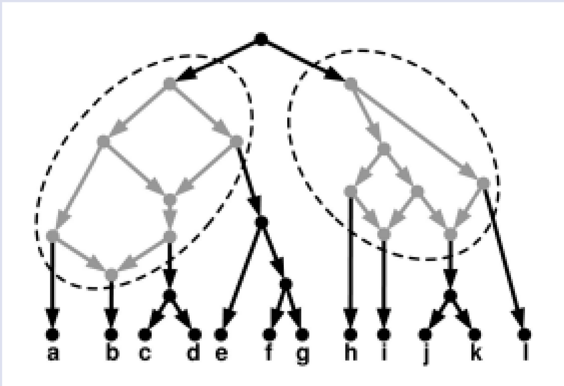

Figure 1: An example of level-2 phylogenetic network which is adapted from [8].

Unrooted and undirected network with distinguished reticulation edges is identified as a semi-directed network. This network is identifiable for Jukes-Cantor large cycle network models, which is a network with a single undirected cylce length at least four [4]. 4 leaf trees are also describes as quartets which are used in tree reconstruction [2]. When I consider the reticulation events, quartets used in network reconstruction recently; for example, SNAQ reconstruct phylogenetic from gene trees [7]. As opposed 4-leaves trees (quartets), I considered the quarnets, which are leaves networks [5]. Huebler et al. constructs level-1 phylogenetic networks from quarnets [6].

In this paper, the work of Huebler et al. by showing that, I can construct in polynomial time a semi-directed level-2 phylogenetic network is extended. An approach for level-2 semi-directed network reconstruction is developed. This method takes the complete set of semi-directed generators displayed by a network and output the associated level-2 network. This methood is based on adding directions to an existing level-2 generator.

In the next section, definitions and notations are introduced. In section 3, a theorem to generate all level-2 generators on five leaves is proved. Finally, section 4 concludes with a discussion about the generators.

Preliminaries

Given an undirected graph G, let V (G) and E (G) be the set of vertices and edges of G, respectively. An undirected graph is biconnected if it contains no vertex whose removal disconnects the graph. A biconnected component, or blob, of an undirected graph is one of its maximal biconnected subgraphs. An unrooted binary phylogenetic network, unrooted network for short, on a set X is an undirected graph where vertices have either degree 3 (internal vertices) or degree 1 (the leaves, labeled univocally by elements in X). An unrooted network is level-k if a tree can be obtained from it by removing at most k edges per blob.

A rooted binary phylogenetic network N {rooted network for short {on a set X is a directed acyclic graph in which exactly one vertex has in-degree 0 and out-degree 2 (the root) and all other vertices have in-degree 1 and out-degree 2 (tree vertices), in-degree 2 and out-degree 1 (hybrid vertices), or in-degree 1 and out-degree 0 (the leaves, distinctly labeled by elements of X). An unrooted network N is level-k if each of its biconnected components contains at most k hybrid vertices.

Denition 1: A semi-directed binary phylogenetic networks {semi-directed networks for short on a set X, which are graphs where:

1. vertices have either degree 3 or degree 1 (the leaves, labeled univocally by elements in X);
2. vertex with degree 3 have either zero (tree vertices) or two (hybrid vertices) incoming directed edges;
3. I can direct all undirected edges to obtain a rooted binary phylogenetic networks without increasing the number of hybrid vertices, possibly having previously subdivided one of the undirected edge with a new node.

An unrooted level-k generator is a biconnected unrooted level-k network, which is not level-(k-1). Rooted/semi-directed level-k generators are defined in a similar way but, for generators, unlike for general rooted/semi-directed binary networks, I allow hybrid vertices to have out-degree 0. In Fig. 3 are depicted all level-0, level-1 and level-2 generators for rooted networks, and in Fig. 24 for semi-directed networks.

For rooted/semi-directed generators, vertices of outdegree 0 and arcs are called its sides; they are empty if no subtree is hanging from them.

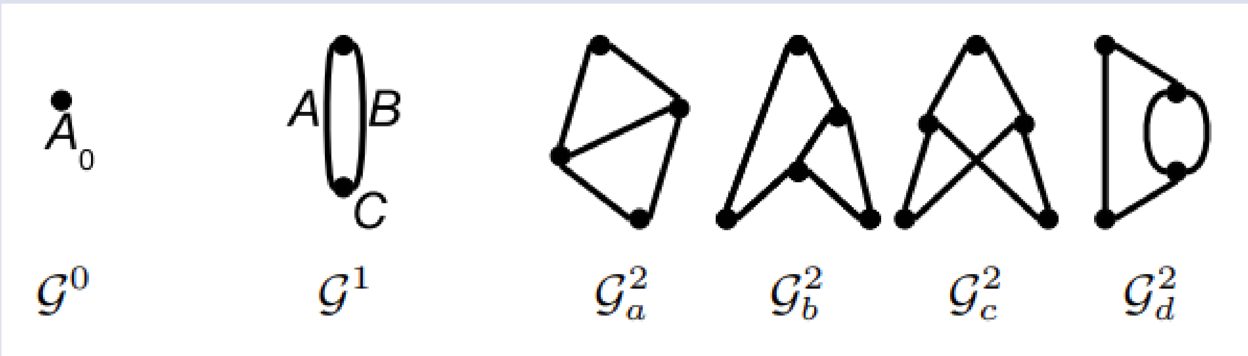

Figure 2: Level-k ≤ 2 generators for rooted networks [1]. G_0 for level-0, G_1 for level-1, and G_a^2, G_b^2, G_c^2, G_d^2 $for\ level-2$.

Unrooted Semi-Directed Level-2 Generators

Level-k generators to build any level-k phylogenetic networks are given in [1]. In this section, we modify these definition and theorems to generate semi-directed level-2 network from generators.

Definition 2: Let G be a level-2 generator. I can adding directed arcs to obtain semi-directed generators given in Figure 3.

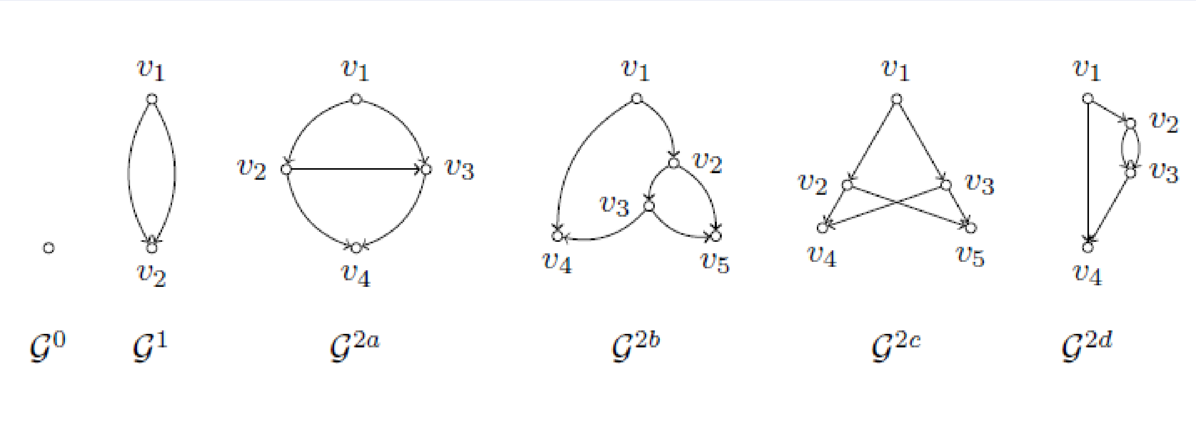

Figure 3: Level-k ≤ 2 semi-directed generators for rooted networks. G_0 for level-0, G_1 for level-1, and G_a^2, G_b^2, G_c^2, G_d^2 $for\ level-2$.

Theorem 2.1 Let G be a level-2 generator. If I adding edges to generators for rooted network to generate semi-directed networks, then the generators for level-2 semi-directed networks are isomorphic to generators for level-2 directed networks.

Proof: In semi-directed networks, it is easy to see that level-0 and level-1 generators coincide with their corresponding rooted ones. Indeed, there is a single biconnected network with no reticulation, which is the trivial network composed of only one vertex i.e., G_0 in Fig. 3. For level-1, there is only one way to have a binary biconnected graph with one hybrid vertex, which is G_1 shown in Fig. 3. In fact, if I make G_1 undirected, I can either make v_1 a hybrid vertex or v_2, obtaining $N^{11}$and N^{12}, which are both isomorphic to G_1 as seen in Figure 4.

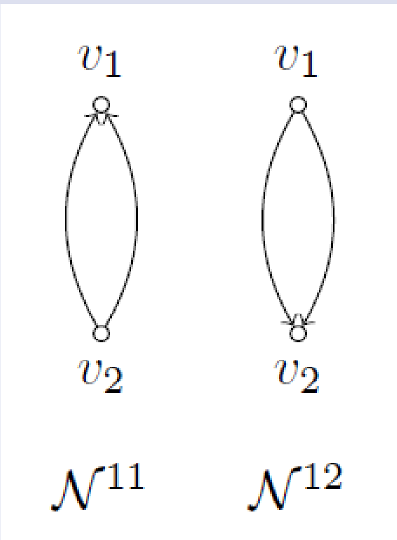

Figure 4: Level-1 semi-directed generators.

For level-2 network, there are four categories of generators, which are obtained respectively from G_a^2, G_b^2, G_c^2, G_d^2 in Fig. 3 by removing the direction of all edges and subsequently choosing a pair of nodes to be the hybrid ones.

Type G^{2a}: Since there are 4 vertices and 2 of them have to be hybrid vertices, we have 6 possible pairs of hybrid vertices: (v_1, v_2), (v_1, v_3), (v_1, v_4), (v_2, v_3), (v_2, v_4), (v_3, v_4).

Case (v_1, v_2): There is just one possible semi-directed network, which is $N^{2a,1}$ in Fig. 6.

Case (v_1, v_3): This case is equivalent to the first one since vertices v_2, and v_3 are interchangeable.

Case (v_1, v_4): There is just one possible semi-directed network, which is $N^{2a,3}$ in Fig. 6. Note that the root position is constrained here: the root needs be placed to subdivide the edge (v_2, v_3) to get a valid network.

Case (v_2, v_3): I need to choose two edges entering in v_2 and two edges entering in v_3 to be the hybrid edges. The nine possibilities are given in Fig. 5. Because of symmetries between v_1 and v_4, and between v_2 and v_3, I have to consider only networks $N^{2a,4}$.

Case (v_2, v_4): This case is equivalent to the first one since vertices v_1 and v_4 are interchangeable.

Case (v_3, v_4): This case is equivalent to the second one since vertices v_1 and v_4 are interchangeable.

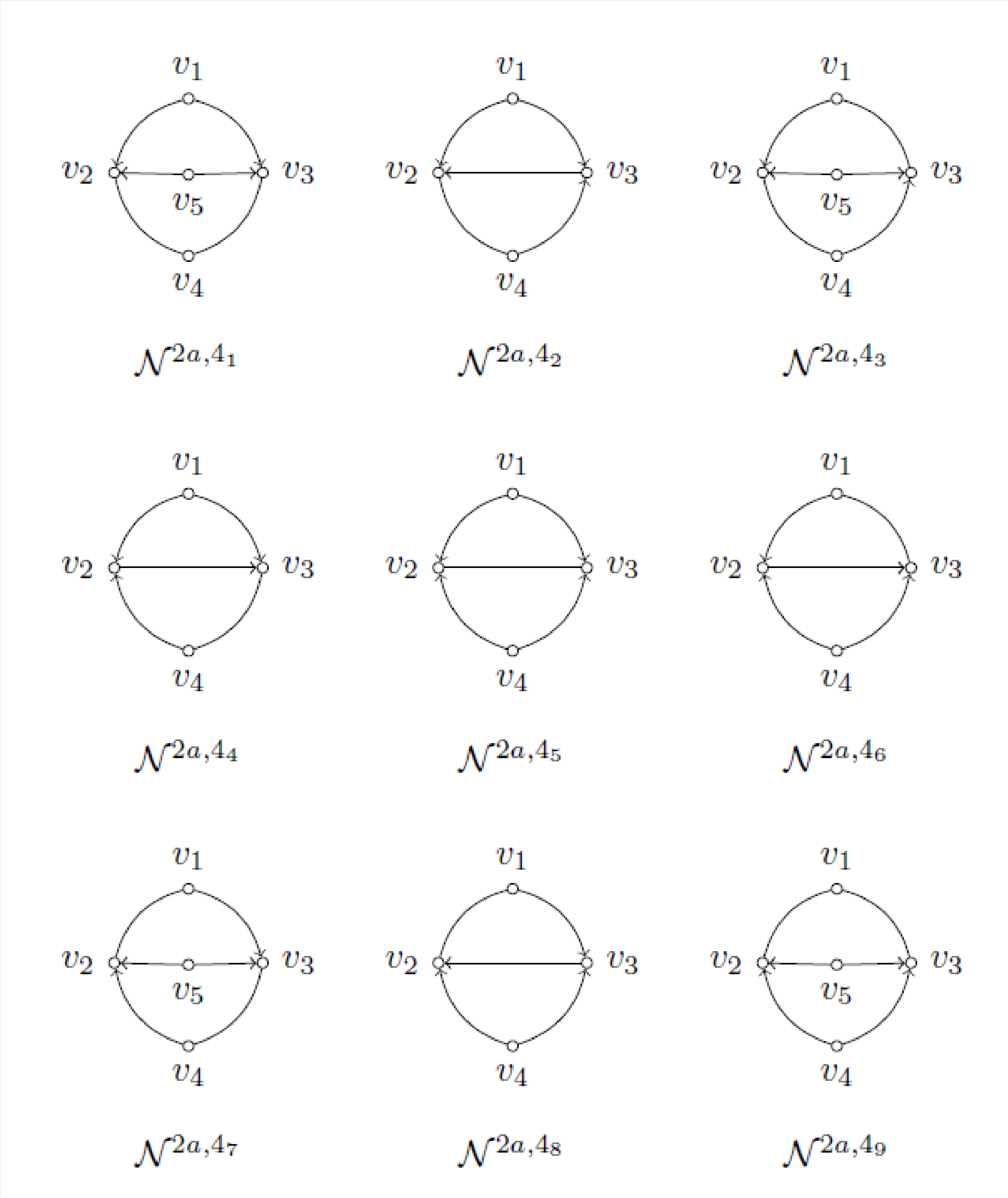

Figure 5: The 9 possible networks obtained by removing edge directions in G^{2a} and choosing v_2 and v_3 as hybrid vertices.

Therefore, I obtain 2 possible generators $N^{2a,1}$ and $N^{2a,3}$ of type (a) corresponding to G^{2a}.

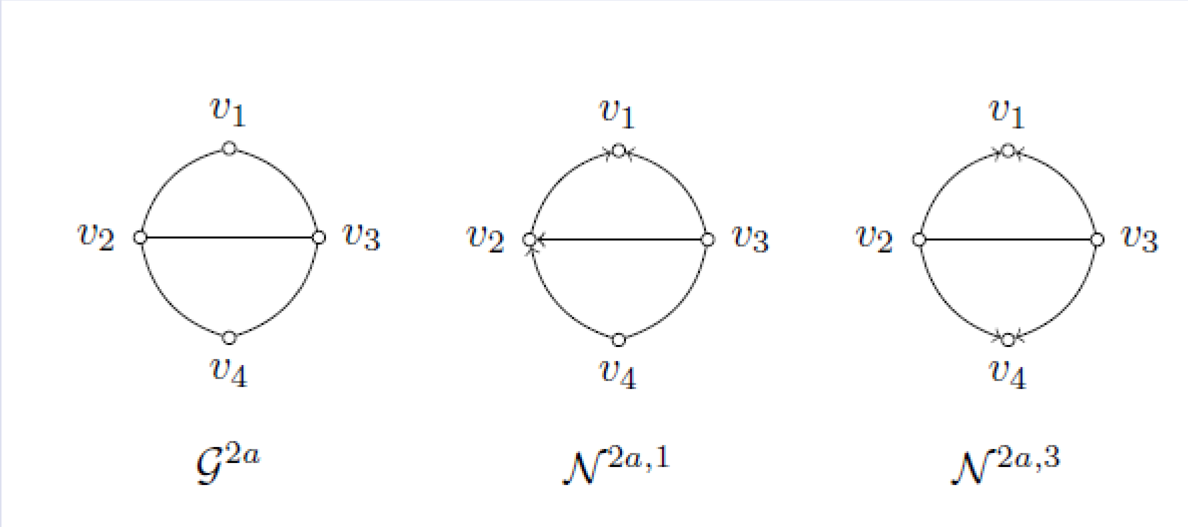

Figure 6: The directed level-2 generator G^{2a} and its corresponding semi-directed generators of type (a) $N^{2a,1}$ and $N^{2a,3}$

Type G^{2b}: Since there are 5 vertices and 2 of them have to be hybrid vertices, I have 10 possible pairs of hybrid vertices: (v_1, v_2), (v_1, v_3), (v_1, v_4), (v_1, v_5), (v_2, v_3), (v_2, v_4), (v_2, v_5), (v_3, v_4), (v_3, v_4), (v_4, v_5).

Case (v_1, v_2): There are three different cases that are shown in Fig. 7 but only $N^{2b,1_1}$ is a proper Level-2 generator.

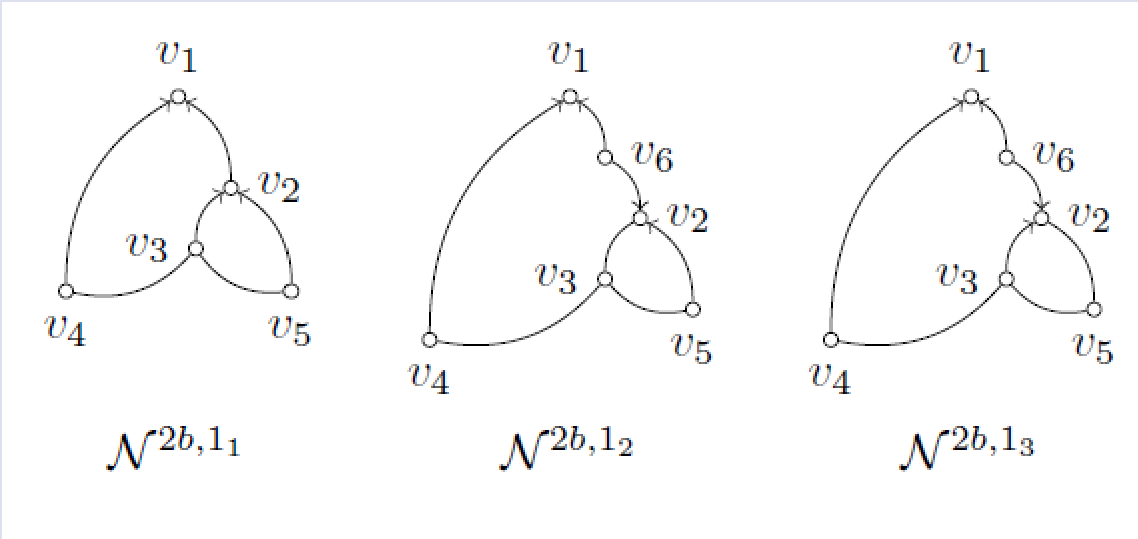

Figure 7: 3 possible $N^{2b,1}$ networks, when v_1 and v_2 are hybrid vertices.

Case (v_1, v_3): There are three different cases that are shown in Fig. 8 but only $N^{2b,1_1}$ is a proper Level-2 generator.

Case (v_1, v_4): In $N^{2b,3}$, I subdivide the edge (v_1, v_4) with a vertex v_6. Assume to have the edge (v_2, v_3), then I also have (v_3, v_5) and (v_2, v_5), and I obtain a cycle (v_2, v_3, v_5). If I have the edge (v_2, v_3), I obtain the reverse cycle. Therefore, $N^{2b,3}$ cannot be a generator because there is no rooted generator that can be obtained from it.

Case (v_1, v_5): There is only one possible network $N^{2b,4}$, which is given in Fig. 14.

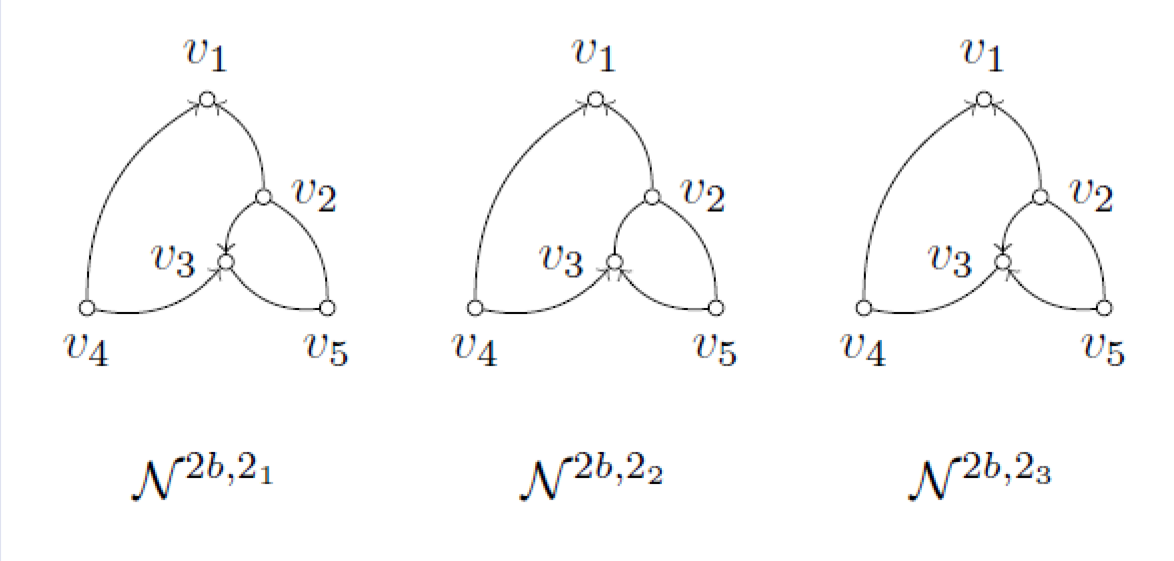

Figure 8: 3 possible $N^{2b,2}$ networks, v_1 and v_3 are hybrid vertices.

Case (v_2, v_3): there are 8 possible networks, see figure 9.

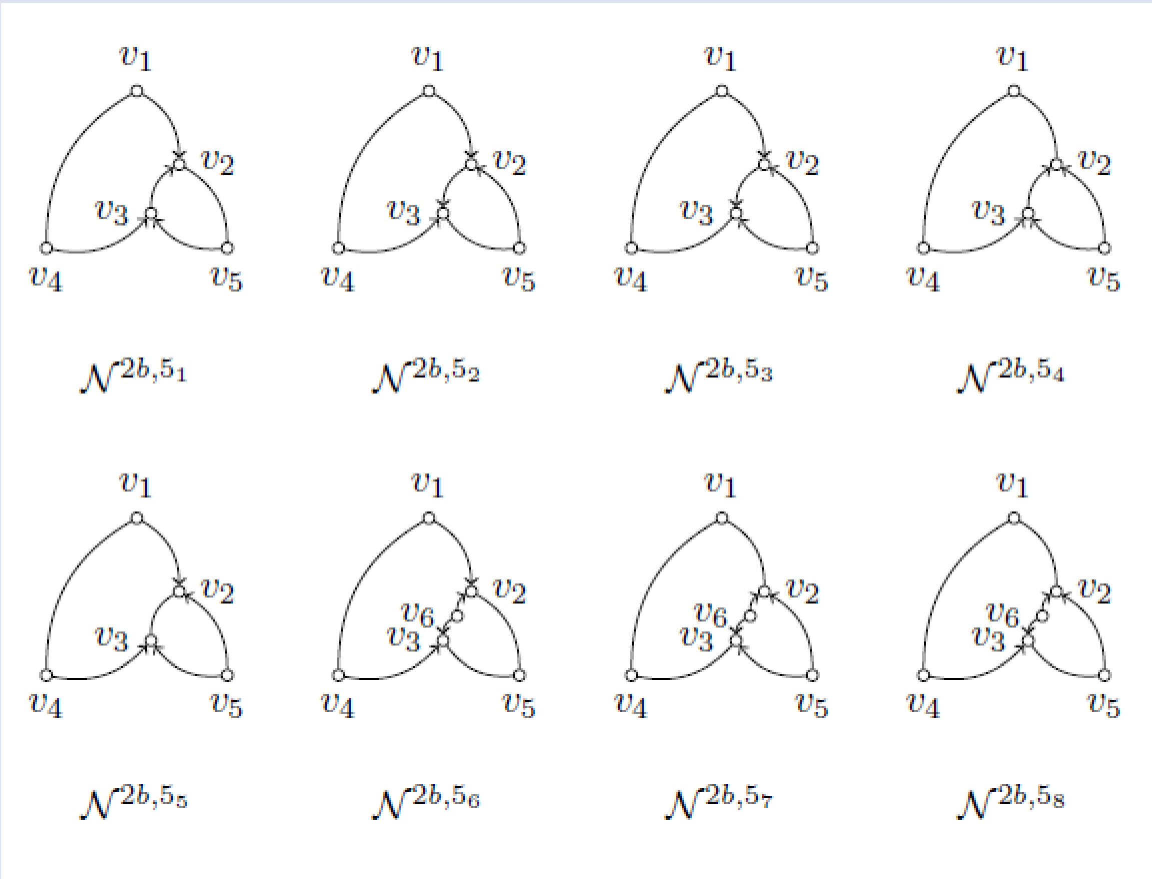

Figure 9: 8 possible $N^{2b,5}$ networks, v_2 and v_3 are hybrid vertices

Case (v_2, v_4): there are 3 possible networks, see figure 10.

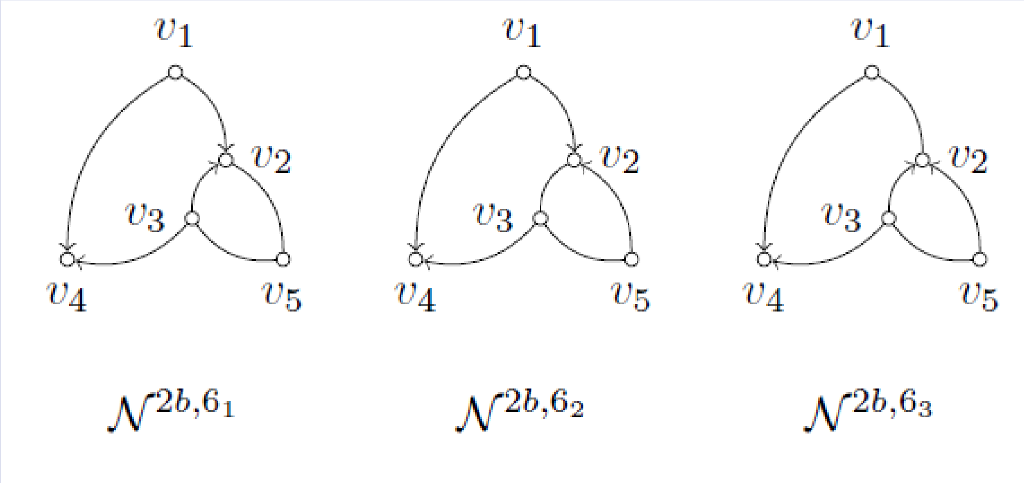

Figure 10: 3 possible $N^{2b,6}$ networks, v_2 and v_4 are hybrid vertices

Case (v_2, v_5): there are 3 possible networks, see figure 11.

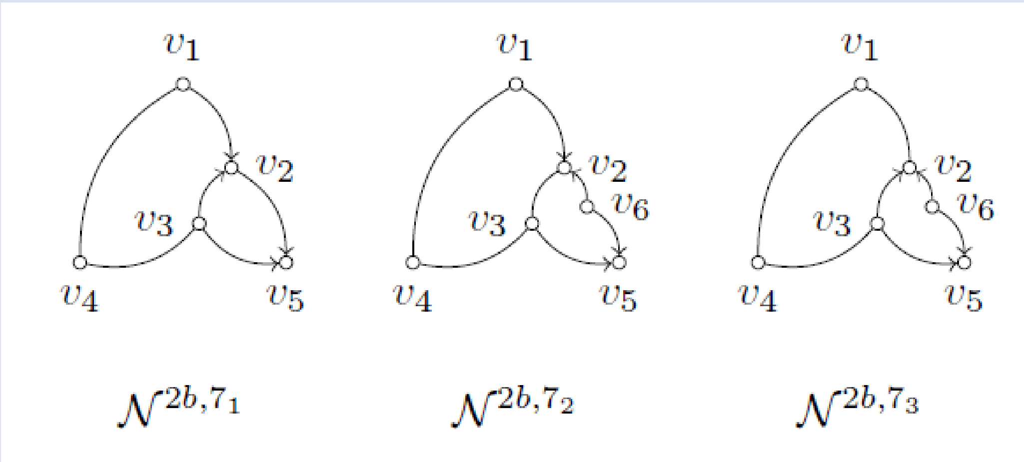

Figure 11: 3 possible $N^{2b,7}$ networks, v_2 and v_5 are hybrid vertices

Case (v_3, v_4): there are three possible networks, see figure 12.

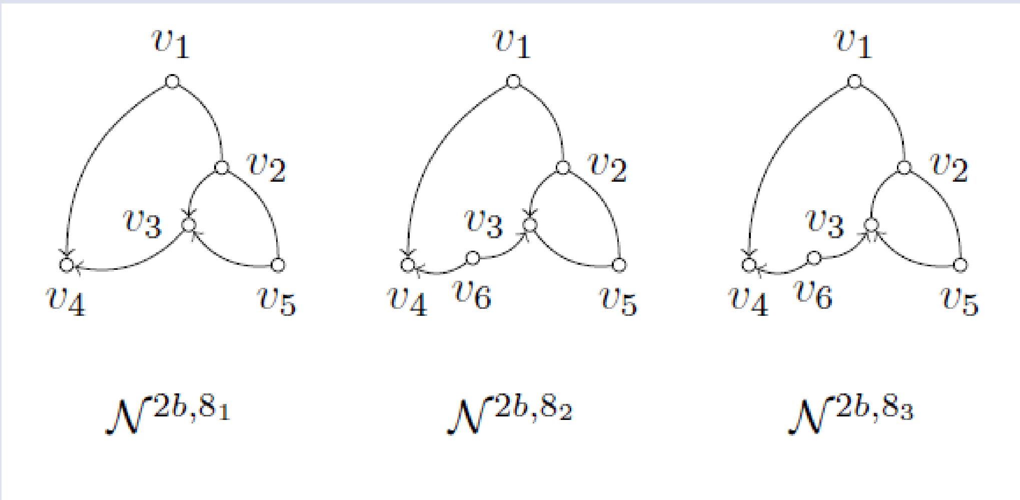

Figure 12: 3 possible $N^{2b,8}$ networks, v_3 and v_4 are hybrid vertices

Case (v_3, v_5): there are 3 possible networks, see figure 13.

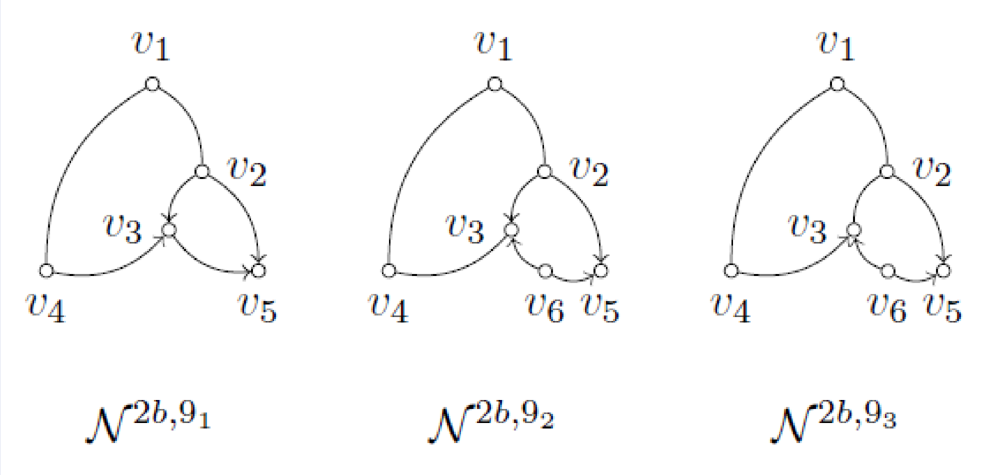

Figure 13: 3 possible $N^{2b,9}$ networks, v_3 and v_5 are hybrid vertices

Case (v_4, v_5): there is only one possible network, see figure 14.

Therefore, I obtain 4 possible generators: $N^{2b,1}$, $N^{2b,2_3}$, $N^{2b,4}$ and $N^{2b,7}$ for a semi-directed network correspondence of Level-2 type (b) generator in the directed network G^{2b} which is given in Fig. 14.

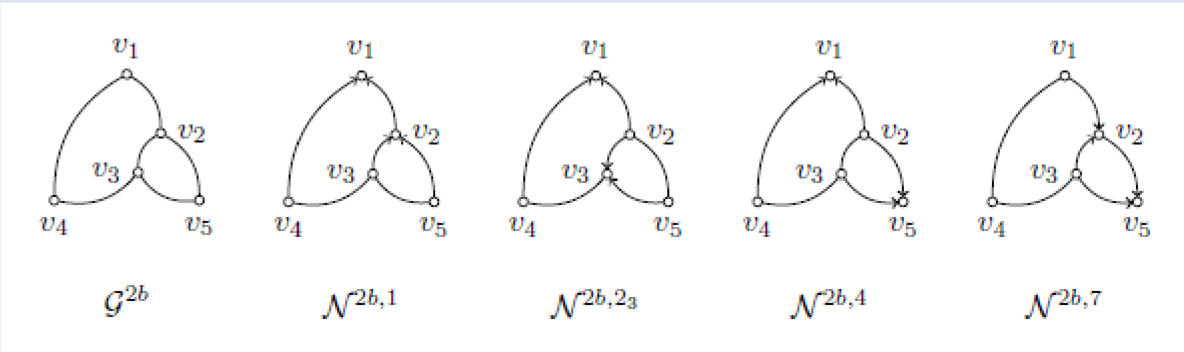

Figure 14: G^{2b} Directed Level-2 type (b) network and $N^{2b,1}$, $N^{2b,2_3}$, $N^{2b,4}$ and $N^{2b,7}$ are generators for semi-directed network Level-2 type (b) network.

For typeG^{2c}, since there are 5 vertices and 2 of them hybrid vertices, C(5, 2) = 10 possible pair of hybrid vertices: (v_1, v_2), (v_1, v_3), (v_1, v_4), (v_1, v_5), (v_2, v_3), (v_2, v_4), (v_2, v_5), (v_3, v_4), (v_3, v_4), (v_4, v_5).

Case (v_1, v_2): there are 3 possible networks, see figure 15. Network $N^{2c,1_1}$ is a possible generator.

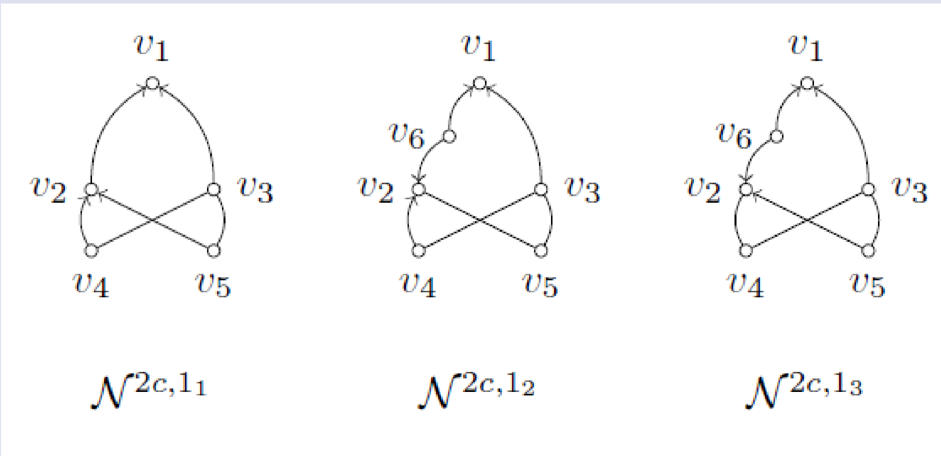

Figure 15: 3 possible $N^{2c,1}$ networks, v_1 and v_2 are hybrid vertices.

Case (v_1, v_3): there are three possible networks, see figure 15. Network $N^{2c,2}$ which are isomorphic to versions of $N^{2c,1}$.

Case (v_1, v_4): there is only one network, and it is a possible generator$N^{2c,3}$, see Fig. 19.

Case (v_1, v_5): there is only one network, and it is a possible generator $N^{2c,4}$, which is isomorphic to$N^{2c,3}$.

Case (v_2, v_3): there are 9 possible networks, see figure 16. Network $N^{2c,5_1}$ and $N^{2c,5_2}$ are possible generators.

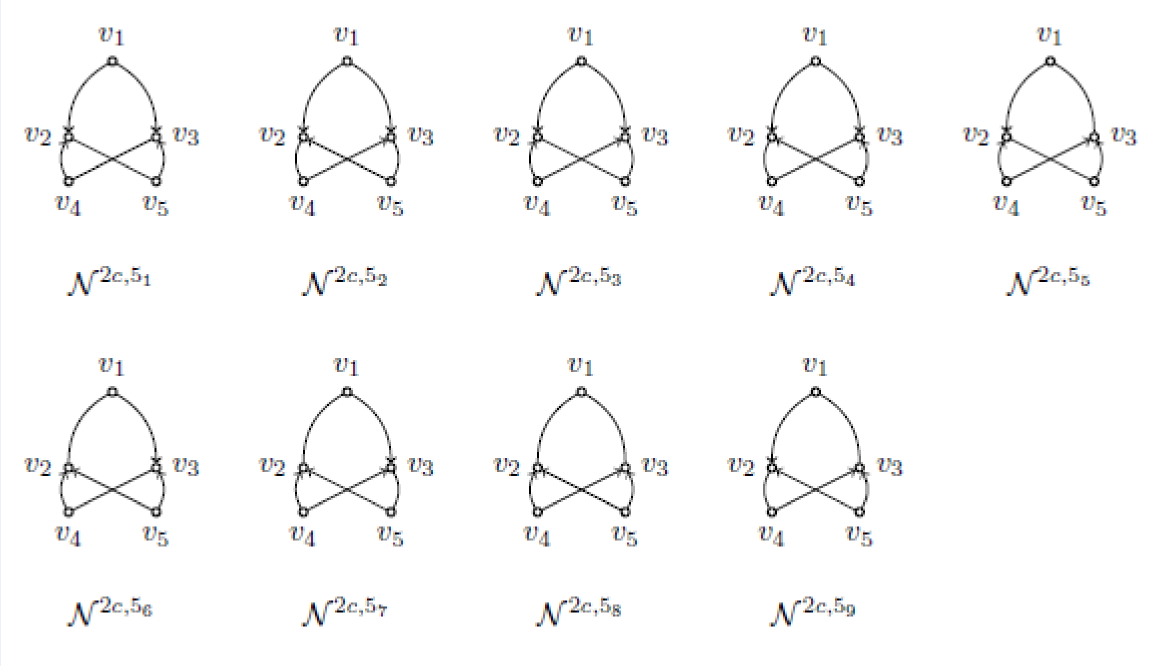

Figure 16: 3 possible $N^{2c,5}$ networks, v_2 and v_3 are hybrid vertices.

Case (v_2, v_4): There are 3 possible networks, see figure 17. Network $N^{2c,6_1}$ is a possible generator.

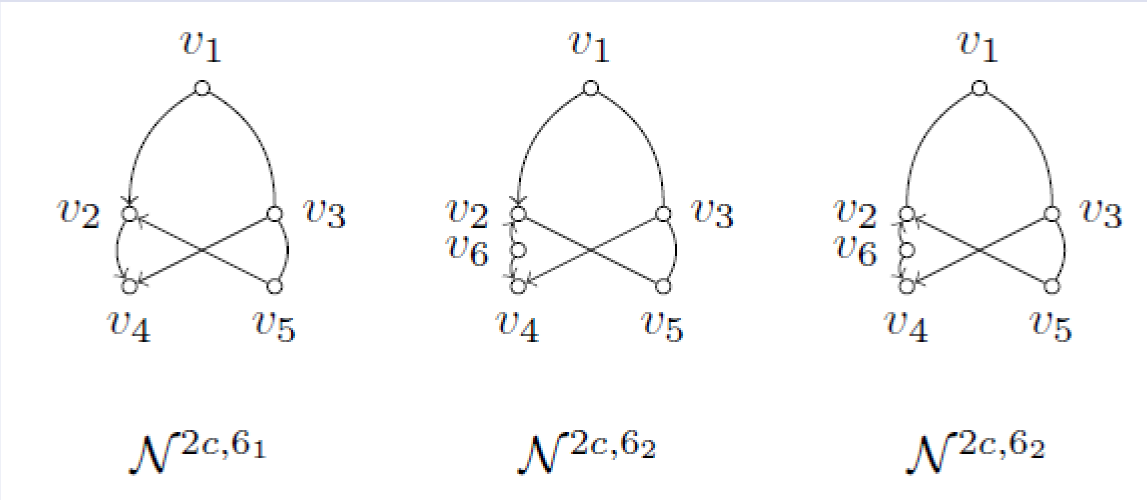

Figure 17: 3 possible $N^{2c,6}$ networks, v_2 and v_4 are hybrid vertices

Case (v_2, v_5): There are 3 possible networks, see figure 18. Network $N^{2c,7_1}$ is a possible generator.

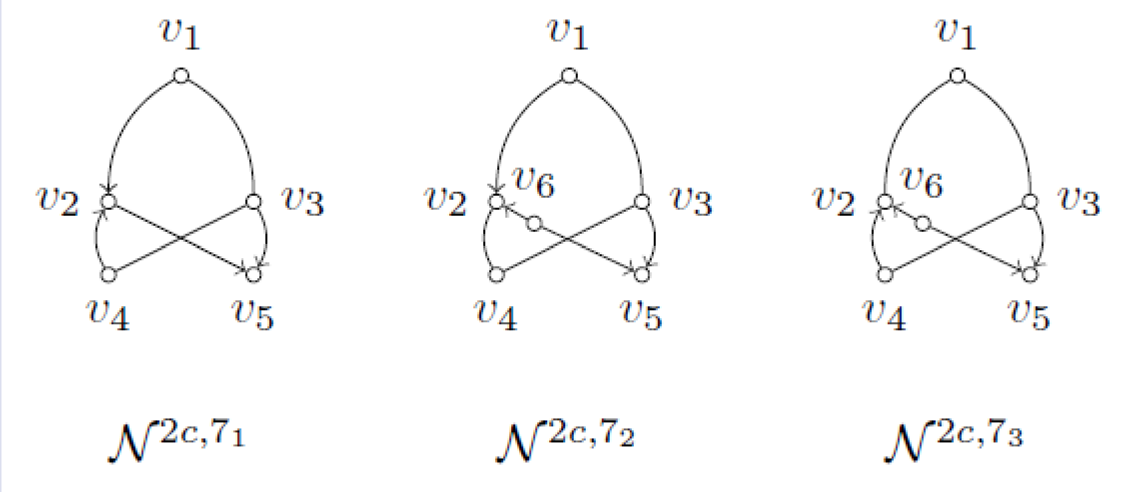

Figure 18: 3 possible $N^{2c,5}$ networks, v_2 and v_3 are hybrid vertices.

Case (v_3, v_4): I obtain 3 version of network $N^{2c,8}$ which are isomorphic to versions of $N^{2c,7}$.
Case (v_3, v_5): I obtain 3 version of network $N^{2c,9}$ which are isomorphic to versions of $N^{2c,6}$.
Case (v_4, v_5): There is only one network, and it is a possible generator $N^{2c,10}$, see Fig. 19.
Therefore, I obtain 7 possible generators $N^{2c,1_1}$,$N^{2c,3}$, $N^{2c,5_1}$, $N^{2c,5_2}$, $N^{2c,6_1}$, $N^{2c,7_1}$ and $N^{2c,10}$ for a semi-directed network correspondence of Level-2 type (c) generator in the directed network G^{2c} which is given in Fig. 19.

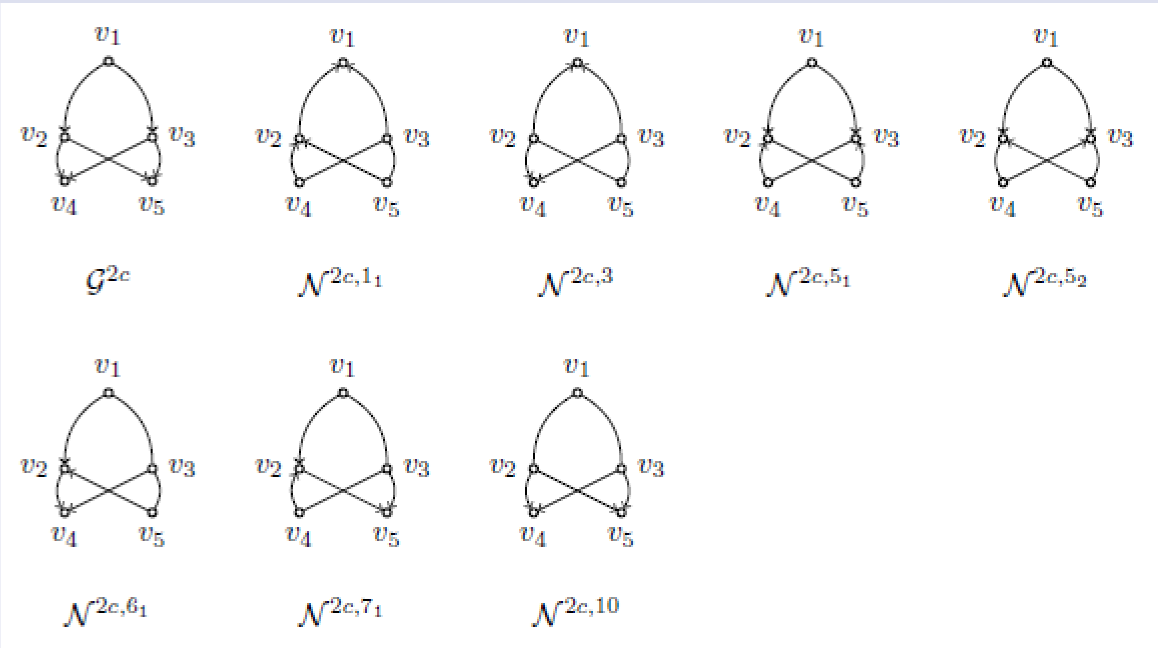

Figure 19: G^{2c} Directed Level-2 type (c) network and $N^{2c,1_1}$, $N^{2c,3}$, $N^{2c,5_1}$, $N^{2c,5_2}$, $N^{2c,6_1}$, $N^{2c,7_1}$ and $N^{2c,10}$are generators for semi-directed network Level-2 type (c) network.

For type G^{2d}, since there are 4 vertices and 2 of them hybrid vertices, C(4, 2) = 6 possible pair of hybrid vertices: (v_1, v_2), (v_1, v_3), (v_1, v_4), (v_2, v_3), (v_2, v_4), (v_3, v_4).
Case (v_1, v_2): There are 3 possible networks, see figure 20. Network $N^{2d,1_1}$ is a possible generator.

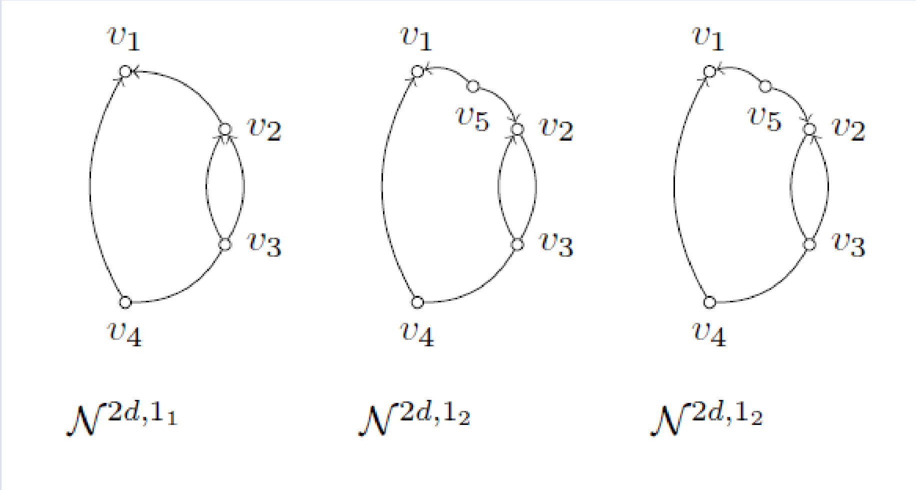

Figure 20: 3 possible $N^{2c,5}$ networks, v_2 and v_3 are hybrid vertices.

Case (v_1, v_3): There are 3 possible networks, see figure 21.

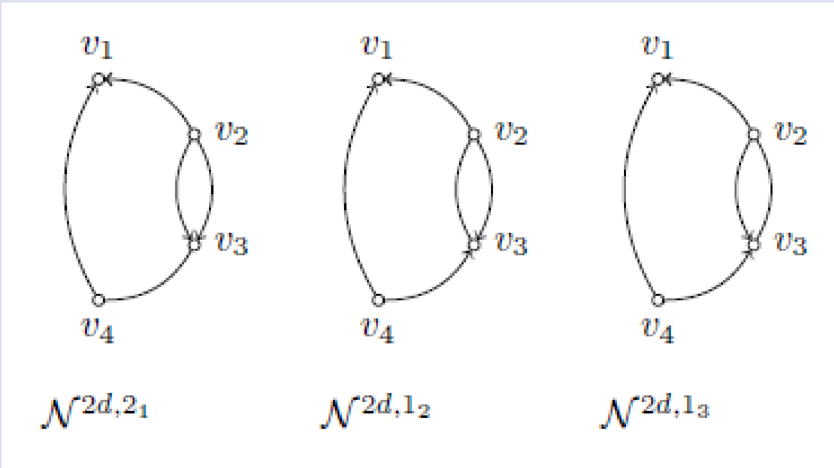

Figure 21: 3 possible $N^{2d,2}$ networks, v_1 and v_3 are hybrid vertices.

Case (v_1, v_4): There is only one network, and it is a possible generator $N^{2d,2_3}$.
Case (v_2, v_3): There are 4 possible networks, see figure 22.

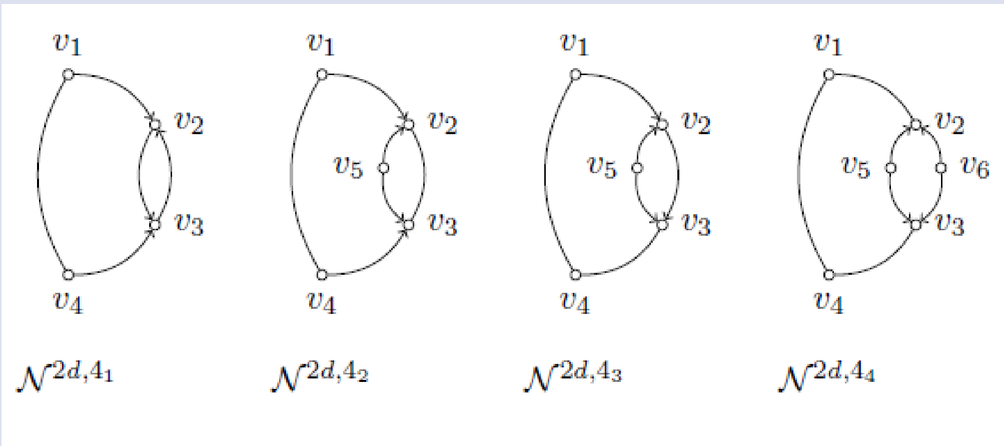

Figure 22: 4 possible $N^{2d,4}$ networks, v_2 and v_3 are hybrid vertices

Case (v_2, v_4): I obtain 3 version of network $N^{2d,5}$ which are isomorphic to versions of $N^{2d,2}$.
Case (v_3, v_4): I obtain 3 version of network $N^{2d,6}$ which are isomorphic to versions of $N^{2d,1}$.
Therefore, I obtain one possible generator $N^{2d,1_1}$ for a semi-directed network correspondence of Level-2 type (d) generator in the directed network G^{2d}, which is given in Fig. 23.

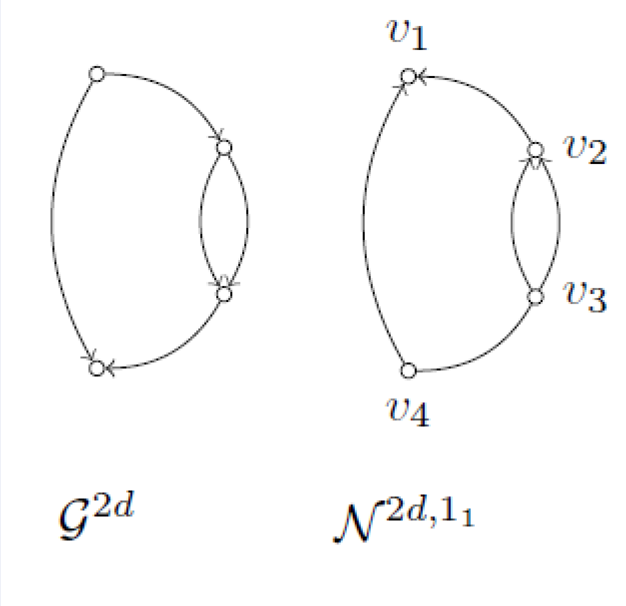

Figure 23: G^{2d} Directed Level-2 type (d) network and $N^{2d,1_1}$ are generators for semi-directed network Level-2 type (d) network.

Results and Discussion

I have shown that by considering level-k generators, I can prove that semi-directed generators isomorphic to level-k generators which is given in Figure 24. For a future work, to show that how I construct a semi-directed level-2 network when a sequence of generators is given.

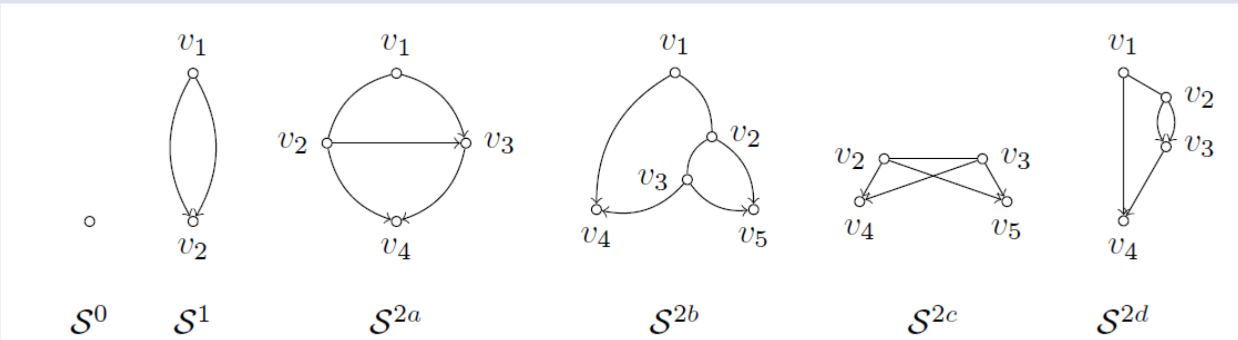

Figure 24: Level-k ≤2 generators for semi-directed networks.

Conflicts of interest

The author state that did not have conflict of interests

References

[1] Gambette P., Berry V., Paul C., The structure of level-k phylogenetic networks, *Annual Symposium on Combinatorial Pattern Matching* Springer, Berlin, Heidelberg, (2009, June). 289-300.

[2] Jansson J., Sung W. K., Inferring a level-1 phylogenetic network from a dense set of rooted triplets, *Theoretical Computer Science*, *363*(1) (2006) 60-68.

[3] Van Iersel L., Keijsper J., Kelk S., Stougie L., Hagen F., Boekhout T., Constructing level-2 phylogenetic networks from triplets, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *6*(4) (2009) 667-681.

[4] Gross E., Long C., Distinguishing phylogenetic networks. *SIAM Journal on Applied Algebra and Geometry*, *2*(1) (2018) 72-93.

[5] Huber K. T., Moulton V., Semple C., Wu T., Quarnet inference rules for level-1 networks, *Bulletin of Mathematical Biology*, *80*(8) (2018) 2137-2153.

[6] Huebler S., Morris R., Rusinko J., Tao Y. Constructing semi-directed level-1 phylogenetic networks from quarnets, (2019) *arXiv preprint arXiv:1910.00048*.

[7] Solís-Lemus C., Ané C., Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting, *PLoS genetics*, 12(3) (2016) e1005896.