Department of Electrical and Computer Engineering

North South University

Senior Design Project

Unsupervised Graph Network Discrimination

Name: Zahin Ahmed **ID # 1711020042**

Name: Mahmud Elahi Akhter **ID # 1721498042**

Faculty Advisor:

Dr. Mohammad Ashrafuzzaman Khan

Assistant Professor

Department of ECE

Summer 2020

LETTER OF TRANSMITTAL

APPROVAL

DECLARATION

ACKNOWLEDGEMENT

ABSTRACT

Table of Contents

CHAPTER 1: INTRODUCTION

1.1 Scope of this Article

In recent years, network analysis techniques have evolved (Chen et al., 2019; Williamson & Tec, 2019) in proportion to the rapid growth of real world networks. Much research has been done on networks such as social networks (Bródka et al., 2011; Culotta & Cutler, 2016), communication networks (Mccallum, 2007), citation networks etc. As time progresses, these networks become bigger and more complex, consequently holding vast amounts of interesting information which can be used for various purposes. Network analysis techniques include, but are not limited to, using random graph models to capture or derive the properties of real world networks (Williamson & Tec, 2019), subgraph isomorphism (Cordella et al., 2004), graph simulation etc. A significant portion of the research conducted on networks has been on centralities (Newman, 2010) to deduce, "which" are the more important nodes in the graph i.e.. central nodes or nodes that propagate most information, and how traffic flows through them. In turn, calculated centralities of various network graphs have been used in many other research fields as well. For example, analyzing social networks, such as tweet classification (Hussain & Islam, 2016), detecting political discussion practices (Miller et al., 2015) and many other such applications (Cohn et al., 2019; Rossman et al., 2010; Yang & Liu, 2008); has involved the generation of multiple centralities. Centralities have also been used in analyzing road traffic networks, such as finding road network patterns (Zhang et al., 2011), tourism management (Lee et al., 2013) etc.. Another important application of centrality measurements is in analyzing biological networks, as can be seen in (Bell et al., 1999; Joyce et al., 2010; Narayanan, 2005; Park & Kim, 2009).

Centralities such as Degree, Closeness, Betweenness, Crossclique, Pagerank, etc. have all been proposed and developed (Crucitti et al., 2006) over the years to answer "Which node is the most influential?" in various different applications (Landherr et al., 2010). "Significance" or "importance" of nodes varies from one context to another. For example, in some cases, it is essential to identify which node(s) propagate more information locally or globally in a graph(Newman, 2010), whereas in other cases, detecting the central node(s) might be of more value (Chiu et al., 2016). Centralities can, therefore, be seen as features/characteristics of a graph; thus, it is possible to discriminate between networks based on their centrality values (Wang & Krim, 2012)

1.2 Organization of article

The article is organized in the following structure. In Chapter 2, we go over previous work that has done on the subject matter. In chapter 3, we describe our selected datasets. Each subchapter describes a different type of dataset i.e. subchapter 3.1 discusses social network datasets and subchapter 3.6 discusses P2P network datasets. In chapter 4 we show the selected centralities and why we selected them. Chapter 5 outlines the models that were constructed for the ablation study. In chapter 6 we show our results and discuss our findings. Subchapters 6.1 and 6.2 are results and subchapter 6.3 is the discussion. We conclude with chapter 7.

1.3 Motivation

In this paper, our main concern is "What information can we extract from a given, unknown graph of a network?" Calculating various centralities from a graph is possible, but this raises the

question of what can be concluded from this information? Is it possible to discriminate between types of networks based on only the centrality values of their nodes? Or, can unsupervised learning techniques applied to unknown graphs produce meaningful context about the network in question? Being able to find meaningful insights about an unidentified network has many practical uses, such as deducing, whether it is a criminal network or not, or if the network was part of a successful political campaign, or whether the network is an ego network of a socially influential person.

CHAPTER 2: LITERATURE REVIEW

Over the years, network centralities have been used for many different purposes that range from traffic network to brain networks. However, there's been very little effort to explore network discrimination through centrality measures. In their work, Wang and Krim("Department of Physics Department of Electrical and Computer Engineering," 2012) showed that discrimination of graph networks are possible through centrality measures. They showed that degree centrality and clustering coefficients were enough to discriminate networks and adding Betweenness Centralities, Eigenvector Centralities or Closeness Centralities degraded the results. However, they only used two small sample datasets for their study. In his study Dwyer(Dwyer et al., 2003) showed that visual analysis to explore and compare the centralities within a given network was possible. In the study the centralities were drawn on a 2D plane that was mapped to 3D plane. Three different methods were employed to this purpose and these were 3D parallel Coordinates-based Comparison, Orbit based comparison and hierarchy based comparison.

Sarrac´en and Rosso (De La Peña Sarracén & Rosso, 2018) used Betweenness centrality to automatically summarize text. In order to do so, they represented the text as an indirected weighted graph where each sentence was represented as a bag of words. They also used similarity/dissimilarity criterion to represent the semantic relation between nodes. Afterwards, a ranking algorithm was used which was based on Betweenness centrality. Wu(Wu et al., 2013) proposed a novel graph clustering algorithm which used Betweenness centrality recursively to create groups of clusters called LEADER which guided the algorithm to cluster the whole network.

Huang et al (Huang et al., 2016) proposed a visual analytics method to explore urban traffic mobility patterns. They used Pagerank and Betweenness centralities to calculate the more central/ important streets. Pagerank detected hub streets and Betweenness detected street/region that acted as back-bone in urban networks.Crucitti (Crucitti et al., 2006) used Closeness, Betweenness, Straightness and information centralities to capture street patterns of different cities of the world. They proposed that a hierarchichal clustering based on distribution of centrality measures are capable of distinguishing different cities to some extent. Zhang (Zhang et al., 2011) used degree, betweennesss and closeness centrality to calculate different road patterns. The aim of their study was to discriminate different road patterns using centralities. The centralities were calculated using a topological network representation of the road networks.

CHAPTER 3: DATASETS

The datasets we have collected belong to the following categories: Social Networks, Citation Networks, Collaboration Networks, Product Co-Purchase Networks, Communication networks, Internet P2P networks, Ground Truth Networks and Signed Networks. The description of each network are given below.

3.1 Social Networks

3.1.1 Ego Networks (Facebook and Twitter)

The Ego networks are essentially networks centered around one particular node. Such networks are formed by taking one node, and finding all the vertices that are directly connected to it while also finding the connections between those vertices. We have used two ego networks, namely the "Ego-Facebook" dataset which contains 4039 nodes and 88233 edges as it combines 10 ego networks, and "Ego-Twitter" dataset which contains 81306 nodes and 2420765 edges. The Facebook ego networks were collected by Facebook users willingly using an app and twitter information was collected using web crawlers.

3.1.2 Page to Page Networks (Facebook)

Another type of network is formed between pages instead of account-holders. The "Gemsec-Facebook" datasets consist of networks of 8 different categories- Artists (50515 nodes and 819306 edges), Athletes (13866 nodes and 86858 edges), Company (14113 nodes and 52310 edges), Government (7057 nodes and 89455 edges), New sites (27917 nodes and 206259 edges), Politicians (5908 nodes and 41729 edges), Public figures (11565 nodes and 67114 edges), and Tv shows (3892 nodes and 17262 edges). Each individual dataset consists of verified Facebook pages of that category, and edges exist between pages that like each other.

3.1.3 Twitch

Twitch is a website on which people can livestream, and it is used mostly by gamers. The "Musae-twitch" datasets consists of 6 networks – DE (9498 nodes and 153138 edges), ENGB (7126 nodes and 35324 edges), ES (4648 nodes and 59382 edges), FR (6549 nodes and 112666 edges), PTBR (1912 nodes and 31299 edges), RU (4385 nodes and 37304 edges) - divided by the languages used by the streamers. The edges represent friendships between the streamers. The data was collected on May 2018.

3.1.4 Last.fm

Last.fm is a website that has been providing music services since 2003. The "Feather-lastfm-social" dataset contains a social network of LastFM users which was collected in March 2020. The nodes represent users residing in Asian countries and edges represent mutual follow relationships. It has 7624 nodes and 27806 edges.

3.2 Collaboration Networks

The Arxiv collaboration networks were collected from the e-print arXiv websites. The edges represent collaborations between authors, so if one author collaborated with another an undirected edge exists between them. The data consists of papers published between January 1993 to April 2003. The “Ca-AstroPH” dataset has 18772 nodes and 396160 edges, and represents collaborations in papers of the Astro Physics category. The “Ca-HepPH” dataset has 12008 nodes and 237010 edges, and represents collaborations in papers of the High Energy Physics - Phenomenology category.

3.3 Citation Networks

The Arxiv citation networks were also collected from the e-print arXiv database. These datasets represent which papers cite each other within this database, and if one paper cites another, a directed edge is drawn from the former to the latter. Information regarding cited papers that do no exist within the database is not present in the networks. The data is collected from papers published between January 1993 and April 2003. The “cit-HepPH” dataset contains 34546 nodes and 421577 edges and consists of papers submitted to the High Energy Physics- Phenomenology category. The “cit-HepTH” dataset contains 27770 nodes and 352807 edges and consists of papers submitted to the High Energy Physics- Theory category.

3.4 Co-purchase Networks

The Amazon product co-purchasing networks were collected by crawling the website and collecting information from the “Customers Who Bought This Item Also Bought” feature. A directed edge exists from one product to another if the former is frequently bought companying the latter. The “Amazon0302” dataset was collected on 2nd March 2003 and has 262111 nodes and 1234876 edges. The “Amazon0601” dataset was collected on 1st June 2003 and has 403394 nodes and 3387388 edges.

3.5 Communication Networks

Communication networks can be constructed by collecting email data from institutes. Such networks are the “Email-EuAll” and “Email-enron” networks. “Email-EuAll” has 265214 nodes and 420045 edges, and was collected from a European institute from October 2003 to May 2005. The nodes represent email address and a directed edge signifies that source node sent at least one email to target node. “Email-enron” has 36692 nodes and 367662 edges and was collected, and made public, from the Enron Corporation when it was being investigated. This network is undirected and contains an edge between nodes if any email was exchanged between them.

3.6 Internet P2P Networks

The hosts in a peep-to-peer sharing system also form networks. Such networks were constructed from the Gnutella file sharing network by taking 9 snapshots across a few days in August 2002.

The ones used in this research are: “P2p-Gnutella04” (10876 nodes and 39994 edges) which was collected on 4th August, “p2p-Gnutella30” (36682 nodes and 88328 edges) which was collected on 30th August, “p2p-Gnutella31” (62586 nodes and 147892 edges) which was taken on 31st August.

3.7 Ground Truth Networks

A network with ground-truth communities available aid in research as they provide a basis for comparison when communities are detected or identified using any algorithm. Such a networks can be constructed from the data available at the DBLP computer science bibliography website. The “Com-DBLP” dataset has 317080 edges and 1049865 nodes and it represents a network of authors. The nodes represent authors and an edge exists between two nodes if they have published at least one paper together.

3.8 Signed Networks

Unlike other social networks where edges between nodes only signify the existence of a relation, but not the nature, signed networks are able to convey both this information. Such a network is the “Soc-sign-Slashdot090221” which has 82140 nodes and 549201 edges and shows whether two connected nodes are friends or foes. However, in this research the sign is not taken into account. The network was collected in February 2009.

The following table shows the details of each dataset, sorted in order of increasing nodes:

Dataset name	Nodes	Edges	Average Degree	Graph Density	Graph Transitivity
musae-Twitch-PTBR	1912	31299	32.73953975	0.008561595	0.130980962
Gemsec-Facebook-Tv shows	3892	17262	8.870503597	0.001139582	0.590643566
Ego-Facebook	4039	88233	43.69051745	0.005408581	0.519189302
musae-Twitch-RU	4385	37304	17.01436716	0.001940065	0.048648019
musae-Twitch-ES	4648	59382	25.55163511	0.00274867	0.084234875
Gemsec-Facebook-Politician	5908	41729	14.12626947	0.00119552	0.301073736
musae-Twitch-FR	6549	112666	34.40708505	0.002626896	0.054128273
Gemsec-Facebook-Government	7057	89455	25.35213263	0.00179624	0.223768822
musae-Twitch-ENGB	7126	35324	9.914117317	0.00069563	0.042433249
Feather-lastfm	7624	27806	7.294333683	0.00047838	0.178622548

musae-Twitch-DE	9498	153138	32.24636766	0.001697535	0.046470889
p2p-Gnutella04	10876	39994	7.354542111	0.000338109	0.005402029
Gemsec-Facebook-Public figure	11565	67114	11.60639862	0.00050179	0.166564488
CA-HepPh	12008	237010	39.47534977	0.00164371	0.659477009
Gemsec-Facebook-Athletes edges	13866	86858	12.52819847	0.00045176	0.129227029
Gemsec-Facebook-Company	14113	52310	7.413023454	0.000262631	0.153198695
CA-AstroPh	18772	396160	42.20754315	0.001124215	0.318001581
Musae-Facebook	22470	171002	15.22047174	0.000338684	0.232321437
Cit-HepTh	27770	352807	25.40921858	0.000457494	0.119569073
Gemsec-Facebook-New sites	27917	206259	14.77658774	0.000264652	0.113984782
Cit-HepPh	34546	421577	24.4067041	0.000353249	0.145656674
p2p-Gnutella30	36682	88328	4.815876997	6.56E-05	0.005163701
Email-Enron	36692	367662	20.04044478	0.00027309	0.085310796
Deezer-RO	41773	125826	6.024274053	7.21E-05	0.075266704
Deezer-HU	47538	222887	9.377214018	9.86E-05	0.092924018
Gemsec-Facebook-Artist edges	50515	819306	32.43812729	0.000321074	0.05349711
Deezer-HR	54573	498202	18.25818628	0.000167282	0.11463005
p2p-Gnutella31	62586	147892	4.726040968	3.78E-05	0.003872019
Ego-Twitter	81306	2420765	59.5470199	0.000366191	0.170572209
soc-sign-Slashdot090221	82140	549201	13.37231556	8.14E-05	0.023617547
Amazon0302	262111	1234876	9.422542358	1.80E-05	0.236082716
Email-EUAll	265214	420045	3.167592963	5.97E-06	0.004106431
com-DBLP	317080	1049865	6.622082755	1.04E-05	0.306377952
Amazon0312	400727	3200440	15.97316877	1.99E-05	0.16048945
Amazon0601	403394	3387388	16.79443918	2.08E-05	0.165622114

CHAPTER 4: CENTRALITIES

4.1 Initial Analysis

Initially, we started off with twenty-six centralities as potential options. These were:

1. Page Rank
2. Closeness Centrality (Latora)
3. Degree Centrality
4. Eigenvector centralities
5. Kleinberg's authority centrality score
6. Kleinberg's hubscore centrality score
7. Betweenness Centrality
8. Centroid value
9. Radiality Centrality
10. ClusterRank
11. DMNC - Density of Maximum Neighborhood Component
12. Clustering coefficient
13. Lobby Index (Centrality)
14. Community Centrality
15. Leverage Centrality
16. Load Centrality
17. Subgraph centrality scores
18. Topological Coefficient
19. Diffusion Degree
20. Eccentricity Centrality
21. Geodesic K-Path Centrality
22. Stress Centrality
23. Information Centrality
24. Markov Centrality
25. Entropy Centrality
26. Local Bridging Centrality

4.2 Final Centralities

Upon seeing the definitions of each centrality and judging whether they are appropriate for our research, we tried calculating each centrality on a medium-sized dataset to see which were calculable within a reasonable amount of time. Finally, we selected 13 final Centralities. These were:

1. **Page Rank**

   PageRank is an eigenvector-based algorithm. The score for a given vertex may be thought of as the fraction of time spent 'visiting' that vertex (measured over all time) in a random walk over the vertices (following outgoing edges from each vertex). PageRank modifies this random walk by adding to the model a probability (specified as 'alpha' in the constructor) of jumping to any vertex. If alpha is 0, this is equivalent to the eigenvector centrality algorithm; if alpha is 1, all vertices will receive the same score (1/|V|). Thus, alpha acts as a sort of score smoothing parameter.

2. **Closeness Centrality (Latora)**

   "Reciprocal of the total distance from a node v to all the other nodes in a network. By definition of shortest-path distances, classic closeness centrality is ill-defined on unconnected networks.

3. **Degree Centrality**

   In graph theory, the degree (or valency) of a vertex of a graph is the number of edges incident to the vertex, with loops counted twice. The maximum degree of a graph G, denoted by $\Delta(G)$, and the minimum degree of a graph, denoted by $\delta(G)$, are the maximum and minimum degree of its vertices. In a regular graph, all degrees are the same, and so we can speak of the degree of the graph. Degree centrality is a local and static metric, since it considers only the directly connected neighbors of a vertex in a static state. Nonetheless, it serves as a useful indicator of the extent of attachment of a vertex to the graph

4. **Eigenvector centralities**

   A measure of importance of nodes in a network using the adjacency and eigenvector matrices. It scores the relative importance of all nodes in the network by weighting connections to highly important nodes more than connections to nodes of low importance. As graph G is undirected and loop-free, the adjacency matrix A is symmetric, and all diagonal entries are 0."

5. **Kleinberg's authority centrality score**

The authority scores of the vertices are defined as the principal eigenvector of t(A)*A, where A is the adjacency matrix of the graph. Obviously, for undirected matrices the adjacency matrix is symmetric and the two scores are the same.

6. Kleinberg's hubscore centrality score

The hub scores of the vertices are defined as the principal eigenvector of A*t(A), where A is the adjacency matrix of the graph.

Obviously, for undirected matrices the adjacency matrix is symmetric and the two scores are the same.

7. Betweenness Centrality

The Betweenness is a node centrality index. It is similar to the stress but provides a more elaborated and informative centrality index. It is calculated considering couples of nodes (v1, v2) and counting the number of shortest paths linking v1 and v2 and passing through a node n. Then, the value is related to the total number of shortest paths linking v1 and v2. Thus, a node can be traversed by only one path linking v1 and v2, but if this path is the only connecting v1 and v2 the node n will score a higher betweenness value (in the stress computation would have had a low score). Thus, a high Betweenness score means that the node, for certain paths, is crucial to maintain node connections. Notably, to know the number of paths for which the node is critical it is necessary to look at the stress. Thus, stress and Betweenness can be used to gain complementary information.

8. DMNC - Density of Maximum Neighborhood Component

For a node v, let N be the node number and E be the edge number of MNC(v), respectively. The score of node v, DMNC(v), is defined to be E/N^{ϵ} for some $1 \leq \epsilon \leq 2$.

We may assume that the MNC has a strong community structure, such as a clique percolation in a random network"

9. Lobby Index (Centrality)

The l-index or lobby index of a node x is the largest integer k such that x has at least k neighbors with a degree of at least k. If x has a high lobby index, then the l-core L(x) (those neighbors which provide the index) has high connectivity (statistically higher than l(x). The authors expect that in the case of social and communication networks (some of which are also based on social networks) the lobby-index is located between the bridgeness, closeness, eigenvector and betweenness centrality

10. Leverage Centrality

Leverage centrality considers the degree of a node relative to its neighbors and operates under the principle that a node in a network is central if its immediate neighbors rely on that node for information. A node with negative leverage centrality is influenced by its neighbors, as the neighbors connect and interact with far more nodes. A node with positive leverage

centrality, on the other hand, influences its neighbors since the neighbors tend to have far fewer connections.

11. Eccentricity Centrality

"The greatest distance between v and any other vertex. The eccentricity of a node in a biological network, for instance a protein signaling network, can be interpreted as the easiness of a protein to be functionally reached by all other proteins in the network. Thus, a protein with high eccentricity, compared to the average eccentricity of the network, will be more easily influenced by the activity of other proteins (the protein is subject to a more stringent or complex regulation) or, conversely could easily influence several other proteins"

12. Information Centrality

"Actor information centrality is a hybrid measure which relates to both path-length indices (e.g., closeness, graph centrality) and to walk-based Eigen measures (e.g., eigenvector centrality, Bonacich power). In particular, the information centrality of a given actor can be understood to be the harmonic average of the "bandwidth" for all paths originating with said individual (where the bandwidth is taken to be inversely related to path length).

13. Local Bridging Centrality

"The local bridging centrality was a variant of global bridging centrality which can be calculated locally requiring only limited local neighborhood graph information. Localized centrality calculations can reduce both communication and computation complexity and there are a myriad of potential applications. It is expected that the concept of localized bridging centrality can improve existing distributed relay or optimization algorithms"

CHAPTER 5: ABLATION STUDY

To understand the effects of different categories of centrality on the results obtained by unsupervised learning algorithm (categories being community centralities and individual node importance centralities), we conducted an Ablation study. We made 9 models where different combinations of centralities were used, and the algorithms were applied on each model separately to see how the results differ and whether a specific combination of centralities aid in proving our hypothesis.

5.1 Model 1

In model 1, all centralities we kept to see the impact of both community and central node driven centralities.

5.2 Model 2

In model 2, we removed only degree. Degree centrality is the most influential centrality in the sense that it is used to calculate many other centralities. Therefore, we wanted to see how removing affects the model

Removed

- × Degree

5.3 Model 3

In model 3, centralities that calculated community were removed along with information centrality. The goal was to see how well the model performs when only centralities that are calculated globally (WITHOUT taking neighborhood/ community into account) are used.

Removed

- × DMNC
- × Leverage
- × Local Bridging
- × Eccentricity
- × Lobby
- × Information

5.4 Model 4

In model 4, centralities that used global information for central node calculation were removed. The idea was to see how the model performed when global centralities were taken out of the equation and centralities that depend on neighborhood/ community were used

Removed

- × Page Rank
- × Closeness
- × Degree
- × Eigenvector
- × Authority
- × Hubscore
- × Betweenness

5.5 Model 5

In model 5, eigenvector, pagerank, hubscore and authority centrality were removed. These centralities calculate importance of a node in similar manner. Therefore, the goal was to see how the impact of "importance" amongst nodes differs from the impact of "distance" between nodes.

Removed

- × Eigenvector
- × Pagerank
- × Hubscore
- × Authority

5.6 Model 6

In model 6, betweenness, eccentricity, closeness and information centrality were removed to see how the impact of "distance" between nodes differs from the impact of "importance" amongst nodes.

Removed

- × Betweenness
- × Eccentricity
- × Closeness
- × Information

5.7 Model 7

Model 7 is mix of both central nodes based centralities and community oriented centralities. In this model leverage, lobby, dmnc, local bridging, information centrality, eigenvector and

closeness were not used. The goal was to see how a model with more global information about "central" nodes with some community information behaves.

Removed

- × Leverage
- × Lobby Index
- × DMNC
- × local bridging-
- × Information centrality
- × Eigenvector
- × Closeness

Used

- ✓ Degree
- ✓ Betweeness
- ✓ Pagerank
- ✓ Hubscore
- ✓ Authority
- ✓ Eccentricity

5.8 Model 8

This is same as model 7. However, here degree, hubscore, authority, betweenness, information, eccentricity and leverage were not used. The purpose was to see the results of equal proportions of community and central node information.

Removed

- × Degree
- × Hubscore
- × Authority
- × Betweenness
- × Information
- × Eccentricity
- × Leverage

Used

- ✓ Page rank
- ✓ Closeness
- ✓ Eigenvector
- ✓ DMNC
- ✓ Local Bridging
- ✓ Lobby

5.9 Model 9

Same as prior 2 models. In model 9, closeness, eigenvector, authority, hubscore, betweenness, leverage and information centrality were unused. The goal was to see how more information about community and some information about central nodes affects the result.

Removed

- × Closeness
- × Eigenvector
- × Authority
- × Hubscore
- × Betweenness
- × Leverage
- × Information Centrality

Used

- ✓ Page Rank,
- ✓ Degree
- ✓ DMNC
- ✓ Lobby Index
- ✓ Eccentricity
- ✓ Local Bridging

CHAPTER 6: RESULTS & ANALYSIS

6.1 P2P-Gnutella04 dataset

6.1.1 Boxplot and Correlation plots

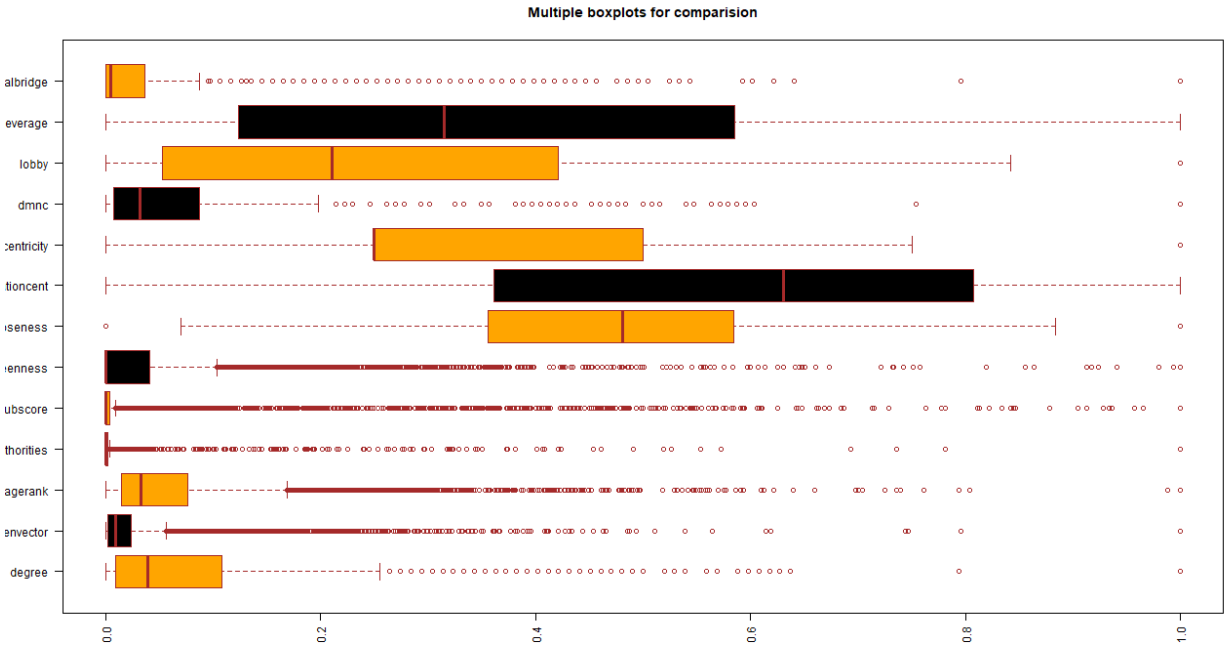


Figure 1 Boxplot

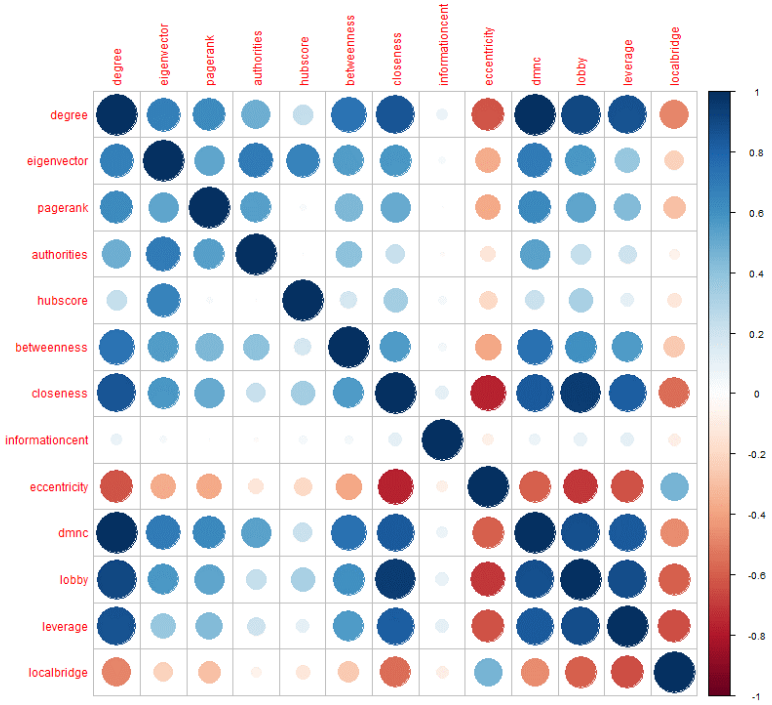


Figure 2 Correlation plot

6.1.2 PCA

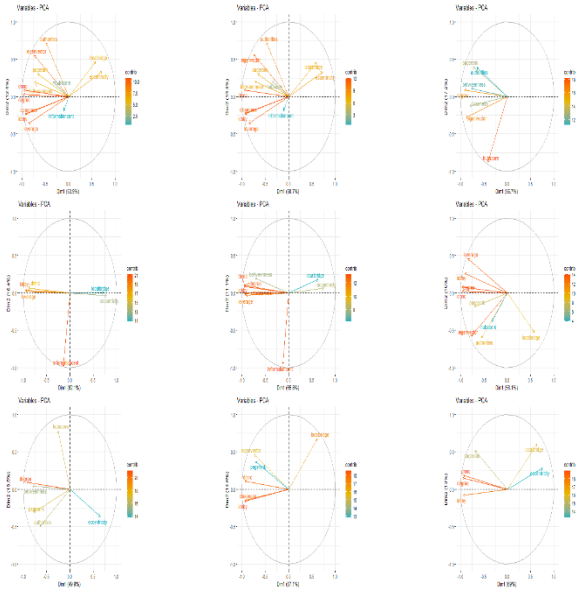


Figure 3 Principal Component Analysis (PCA)

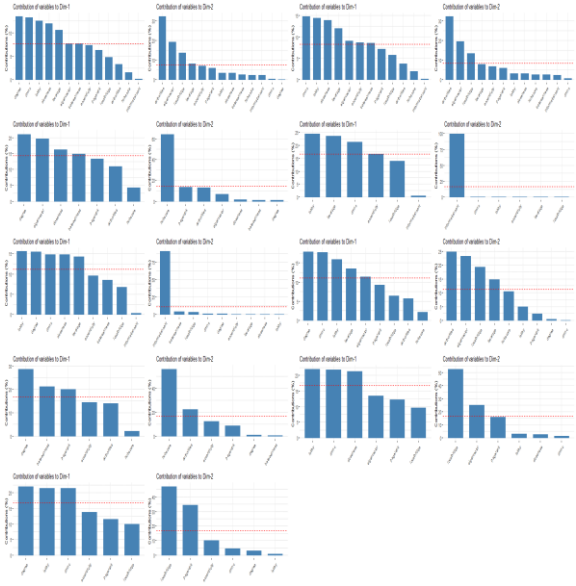


Figure 4 Dimension Contribution

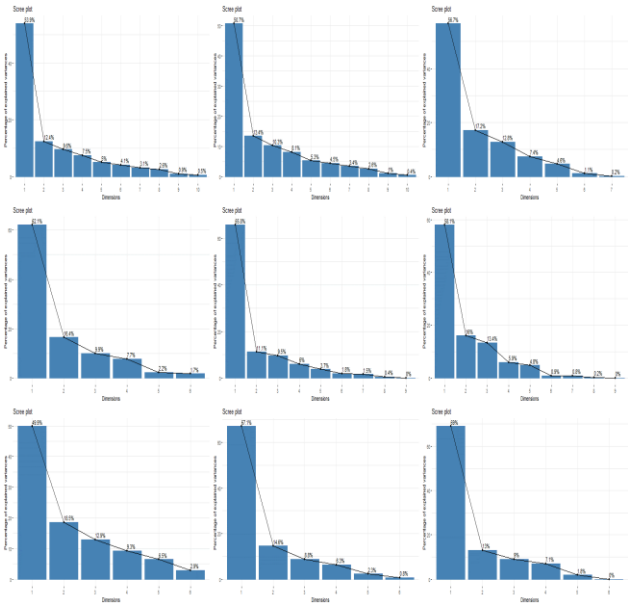


Figure 5 Scree plot

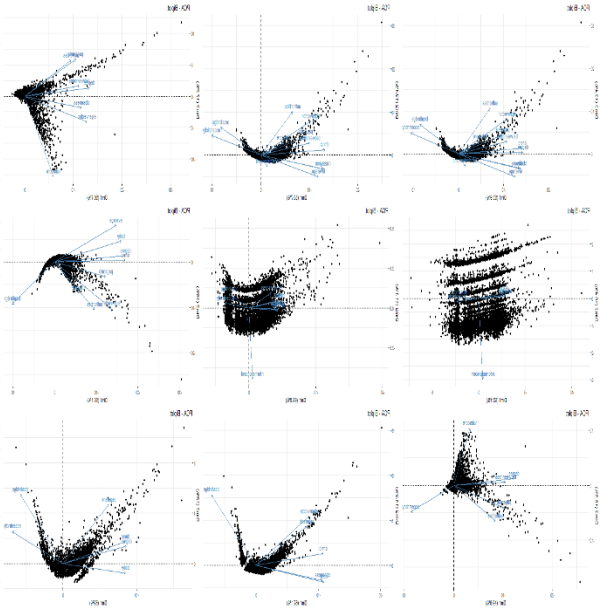


Figure 6 Biplot

6.1.3 t-SNE and K-means on t-SNE

t-SNE Model 1:

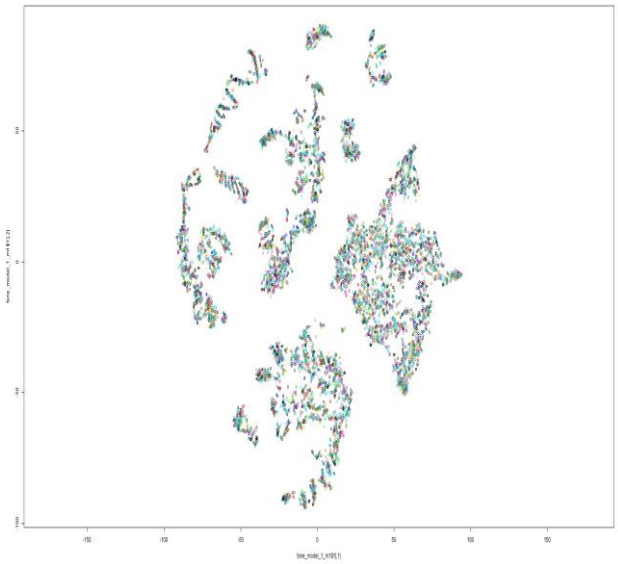


Figure 7 t-SNE on Model 1

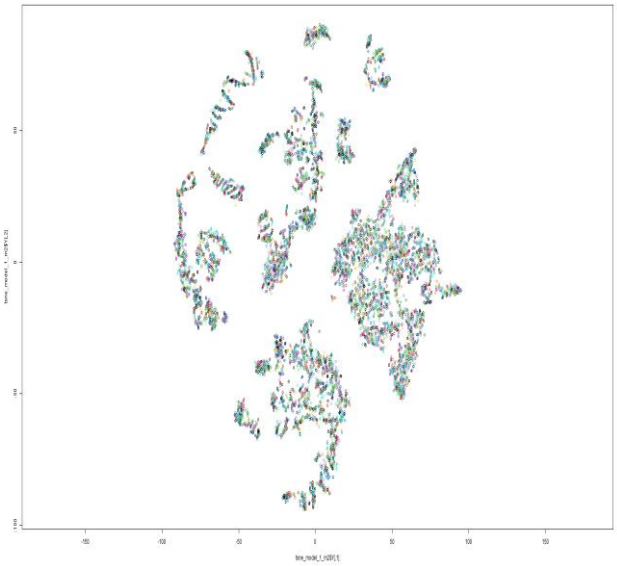


Figure 8 t-SNE on Model 2

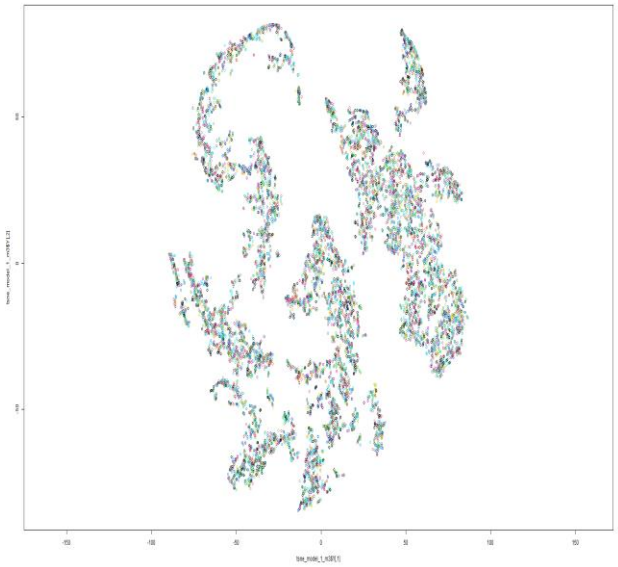


Figure 9 t-SNE on Model 3

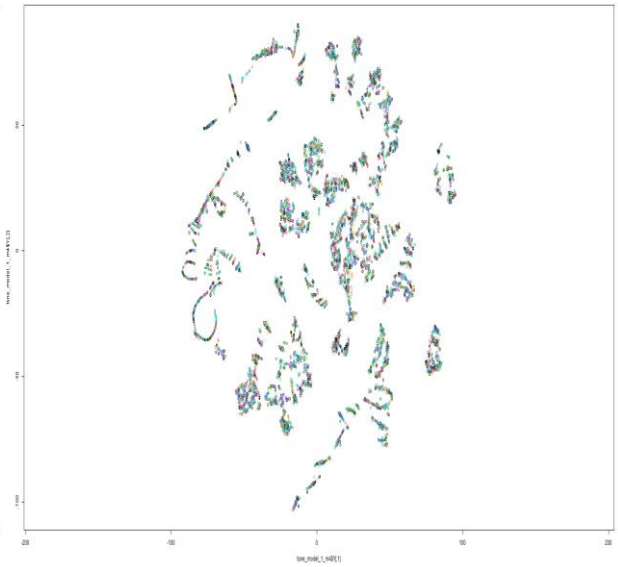


Figure 10 t-SNE on Model 4

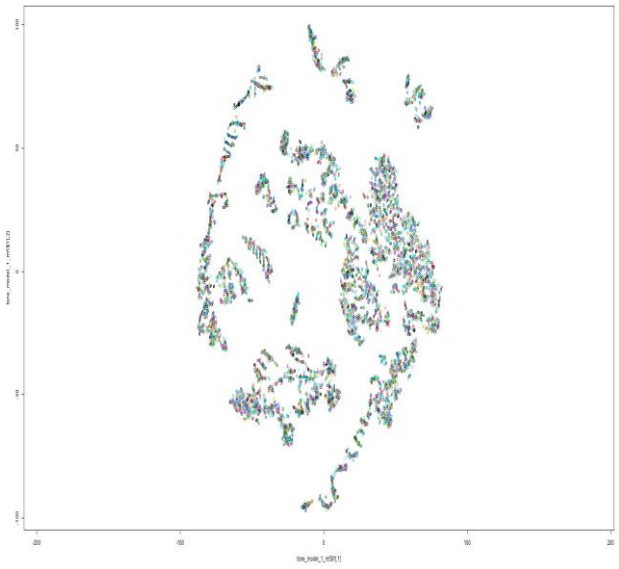


Figure 11 t-SNE on Model 5

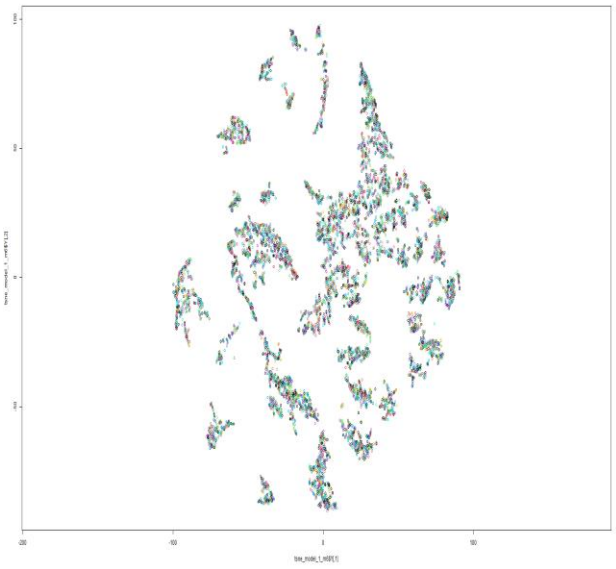


Figure 12 t-SNE on Model 6

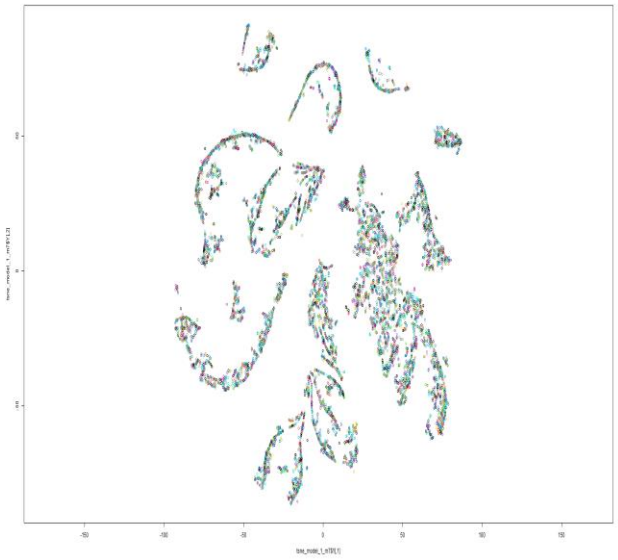


Figure 13 t-SNE on Model 7

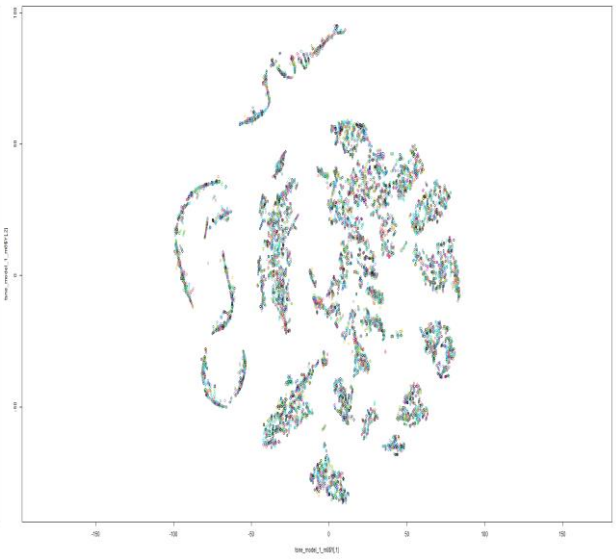
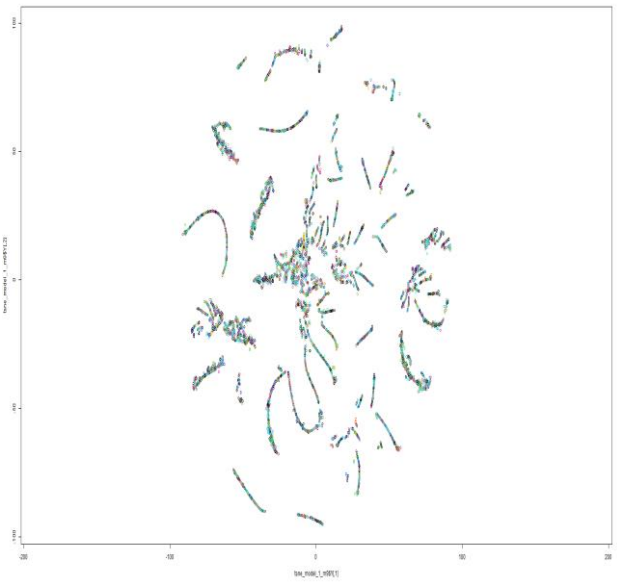


Figure 14 t-SNE on Model 8



t-SNE Model 2:

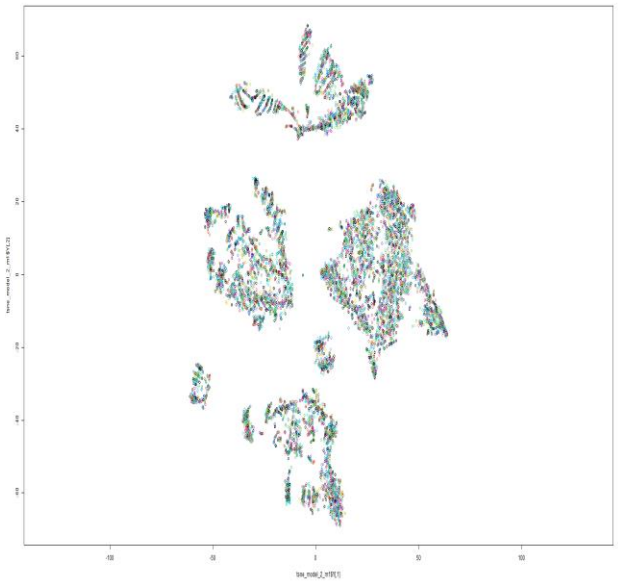


Figure 15 t-SNE on Model 1

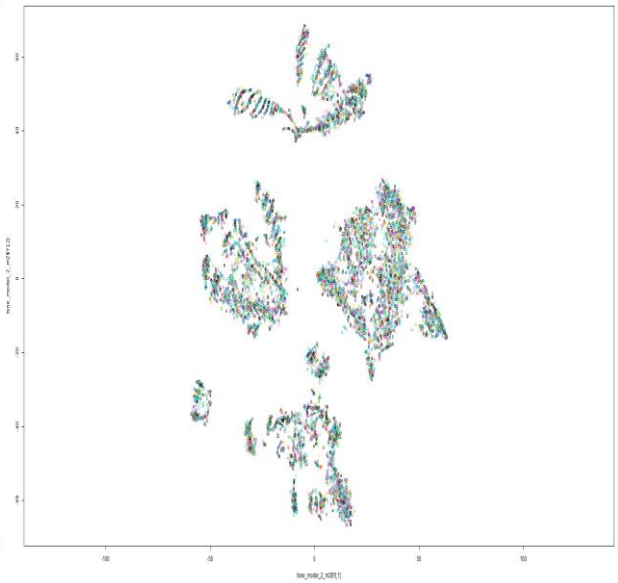


Figure 16 t-SNE on Model 2

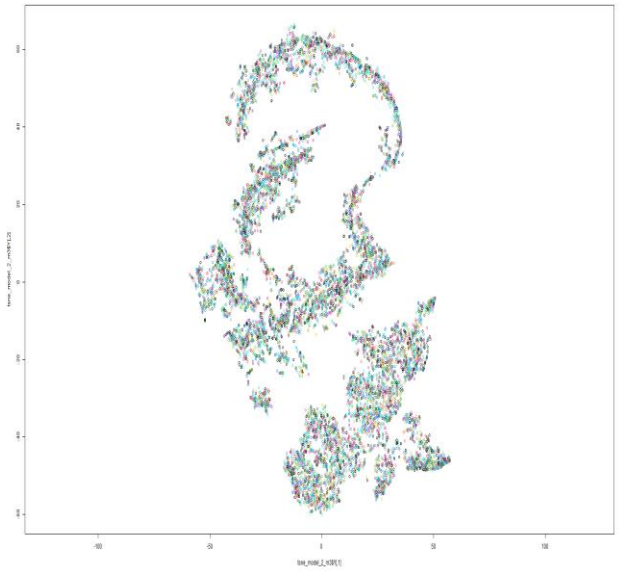


Figure 18 t-SNE on Model 3

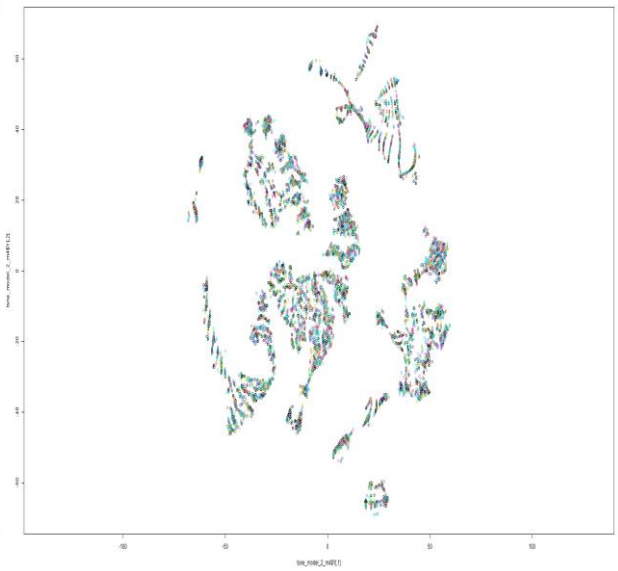


Figure 19 t-SNE on Model 4

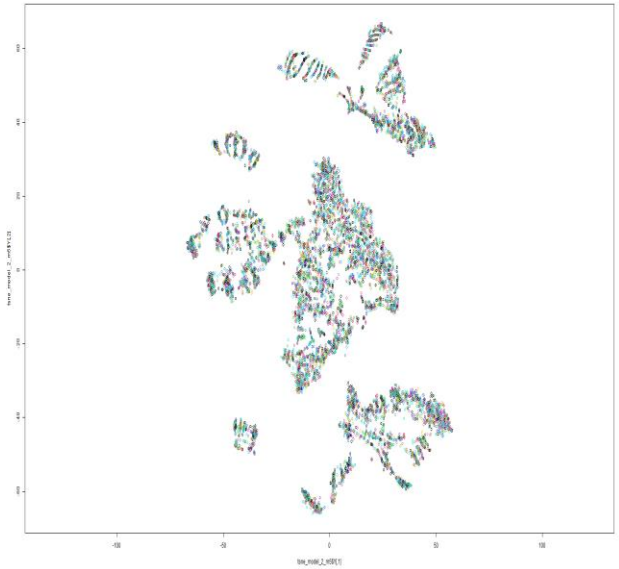


Figure 20 t-SNE on Model 5

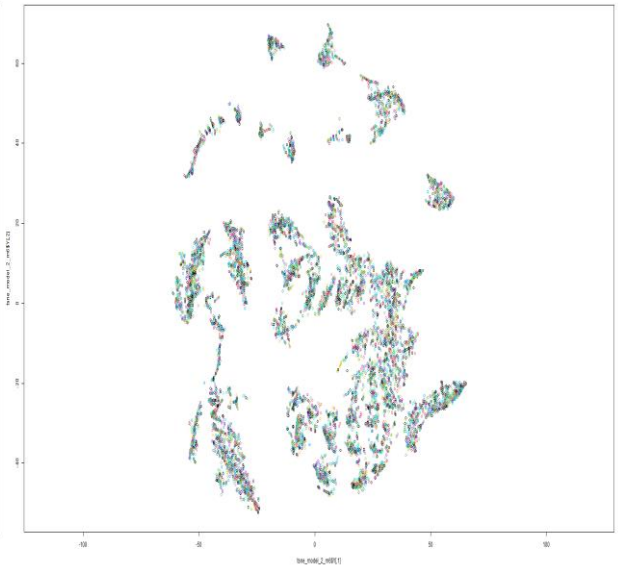


Figure 21 t-SNE on Model 6

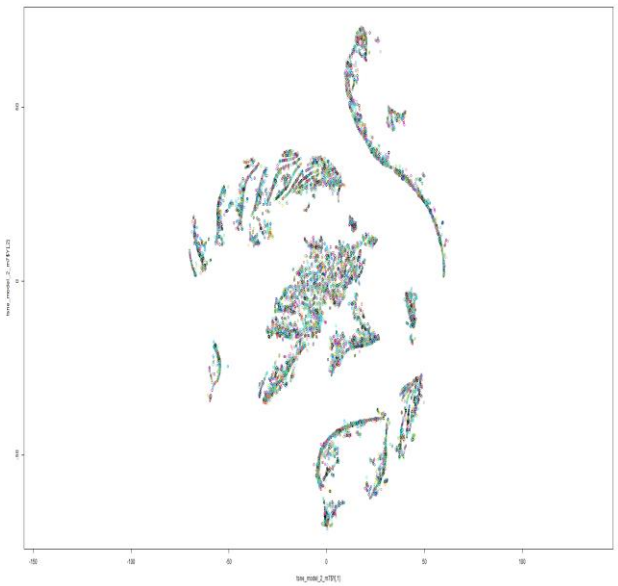


Figure 22 t-SNE on Model 7

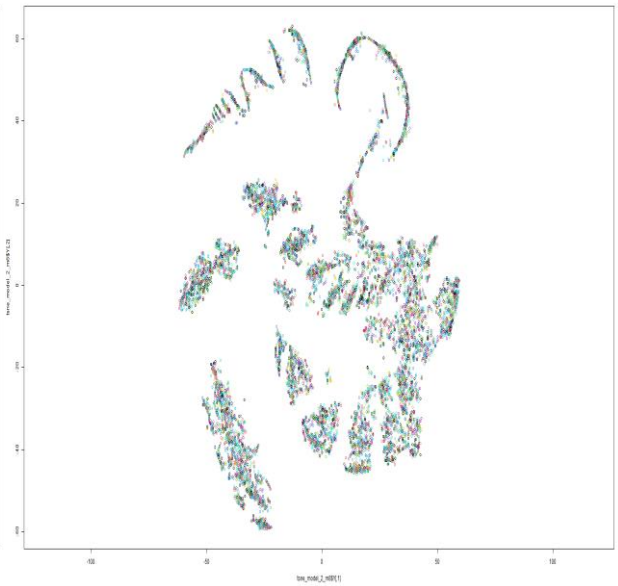


Figure 23 t-SNE on Model 8

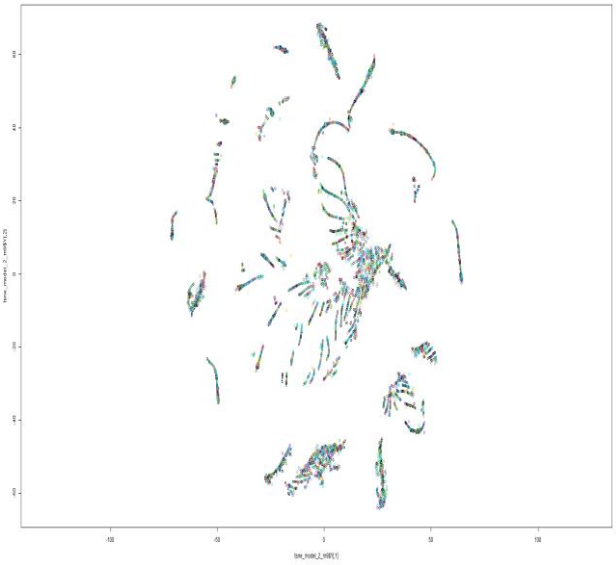


Figure 24 t-SNE on Model 9

t-SNE Model 3:

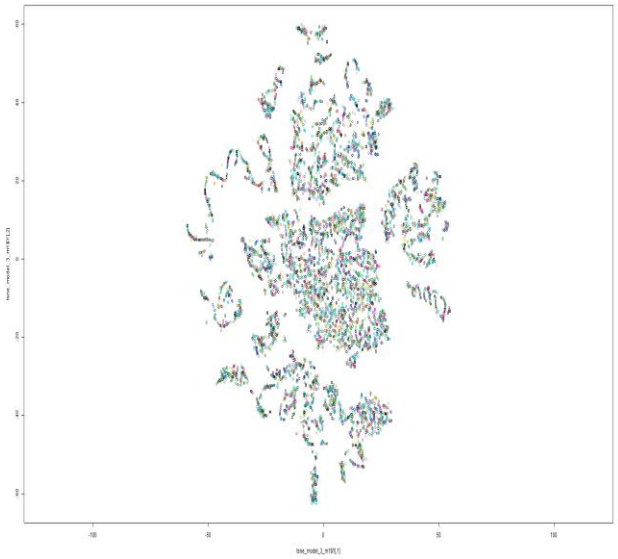


Figure 25 t-SNE on Model 1

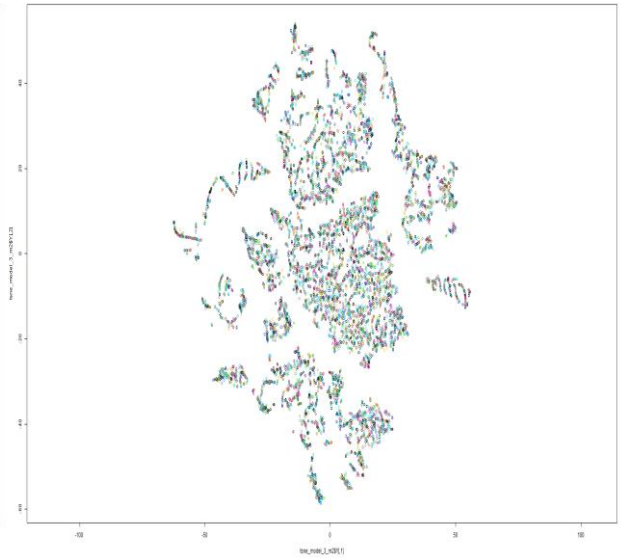


Figure 26 t-SNE on Model 2

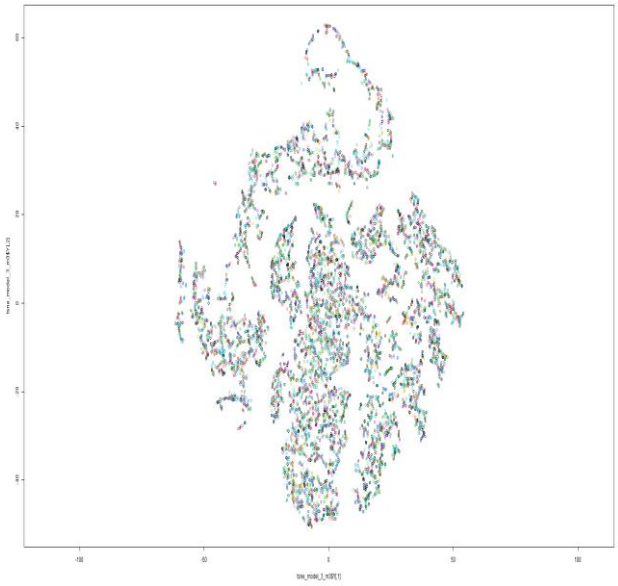


Figure 27 t-SNE on Model 3

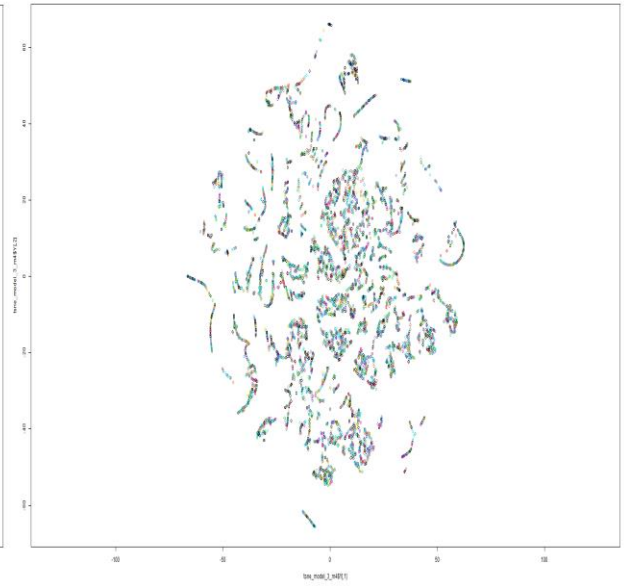


Figure 28 t-SNE on Model 4

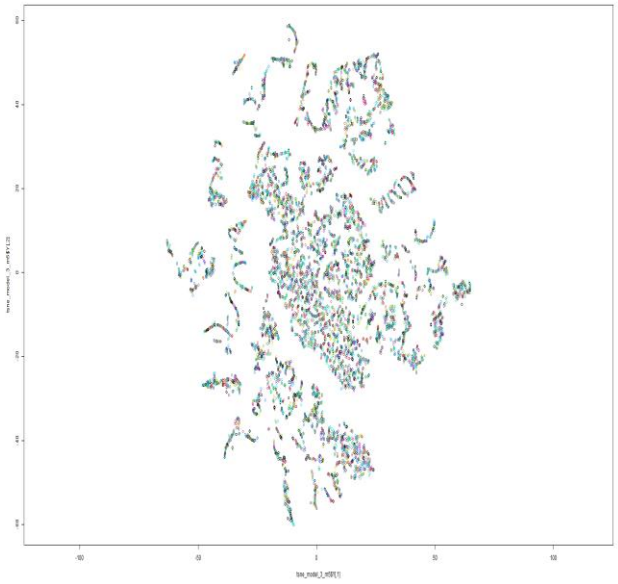


Figure 29 t-SNE on Model 5

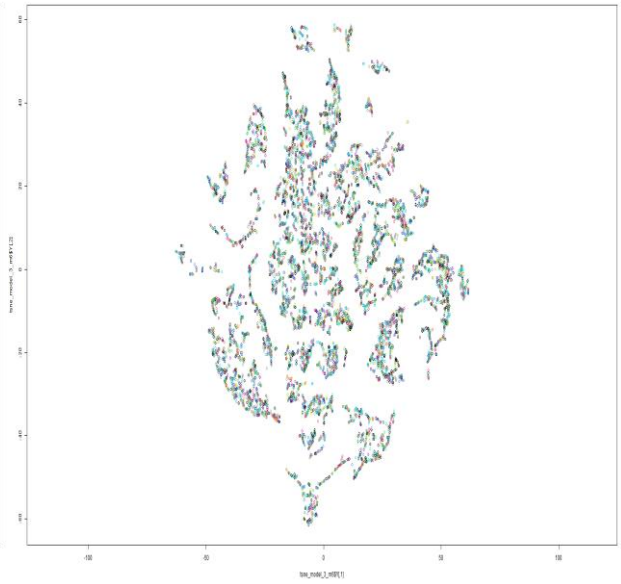


Figure 30 t-SNE on Model 6

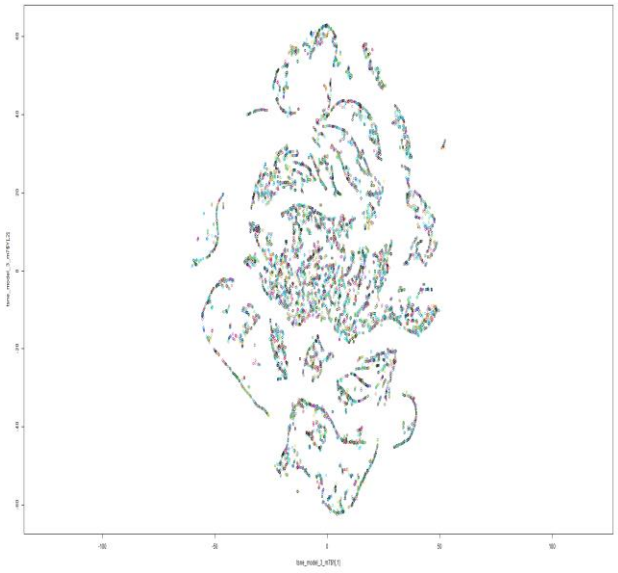


Figure 31 t-SNE on Model 7

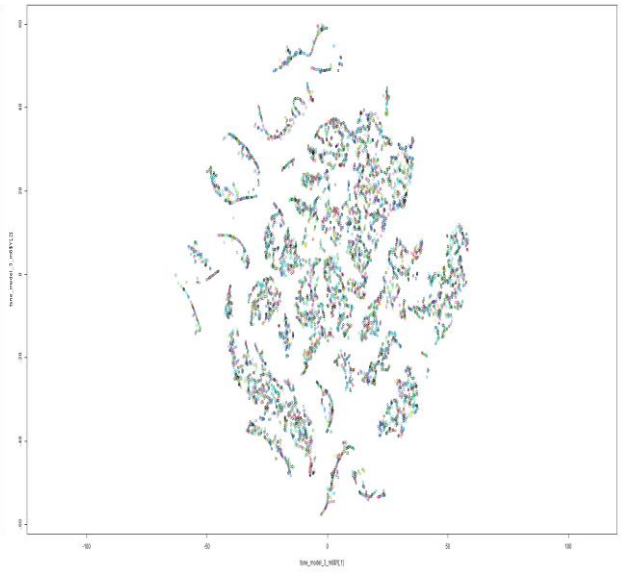


Figure 32 t-SNE on Model 8

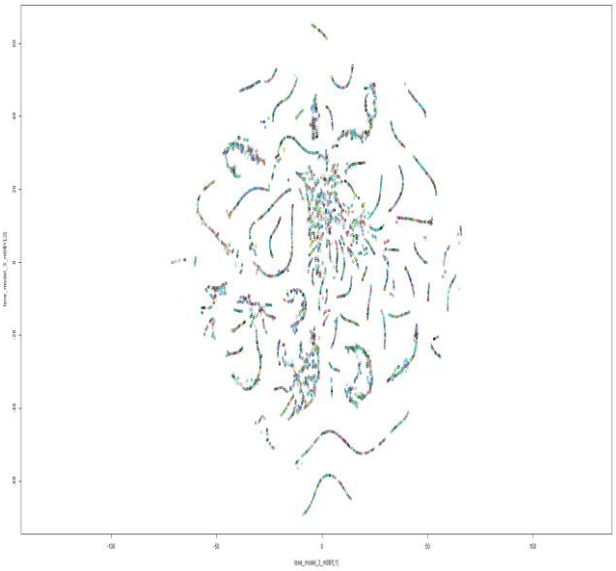


Figure 33 t-SNE on Model 9

t-SNE Model 4:

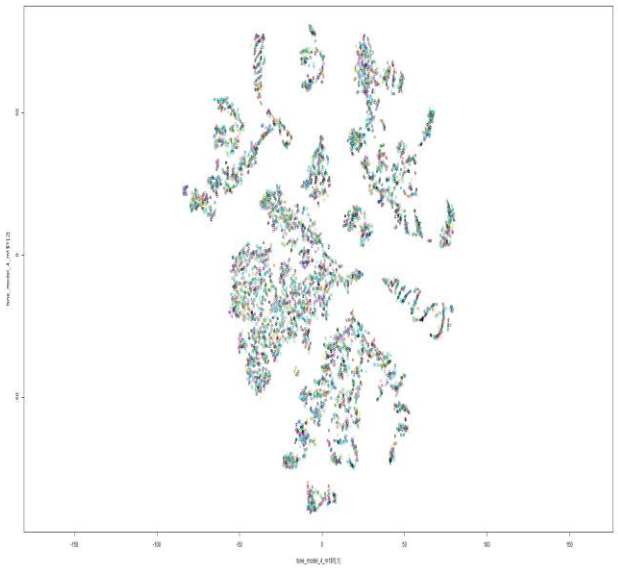


Figure 34 t-SNE on Model 1

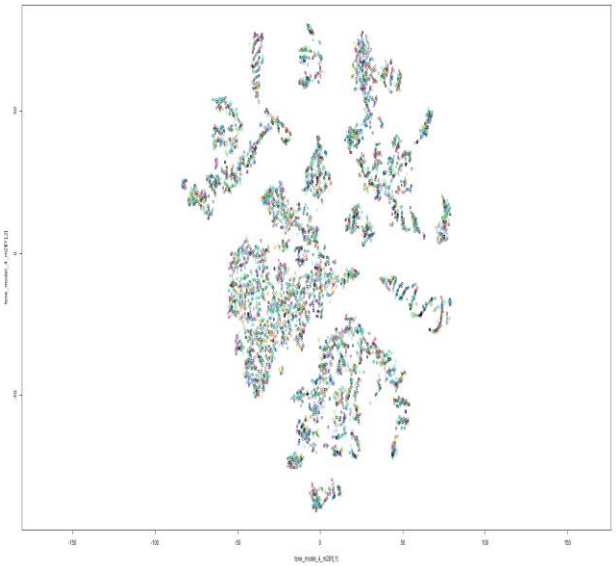


Figure 35 t-SNE on Model 2

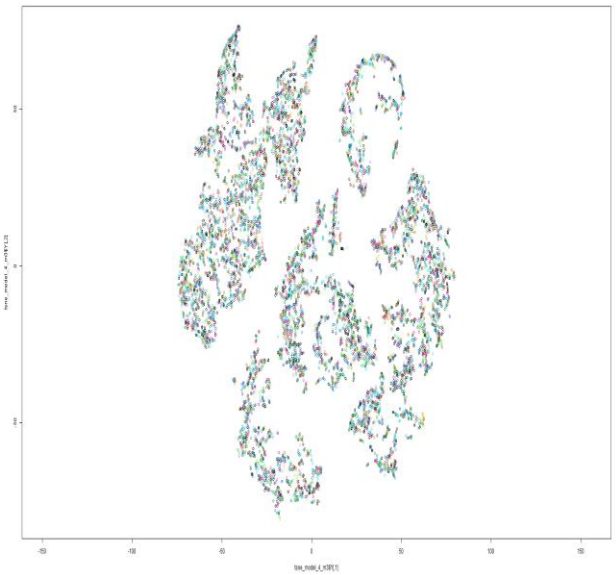


Figure 36 t-SNE on Model 3

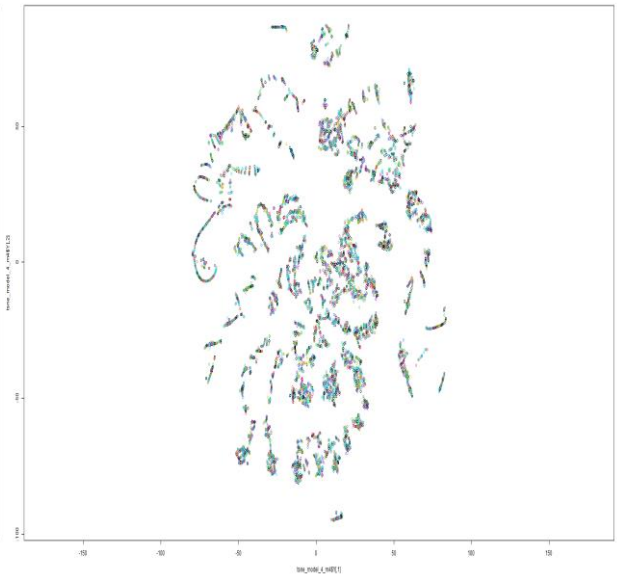


Figure 37 t-SNE on Model 4

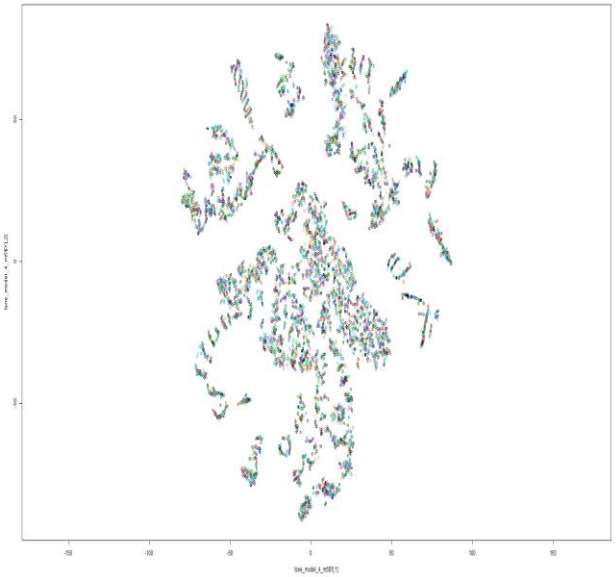


Figure 38 t-SNE on Model 5

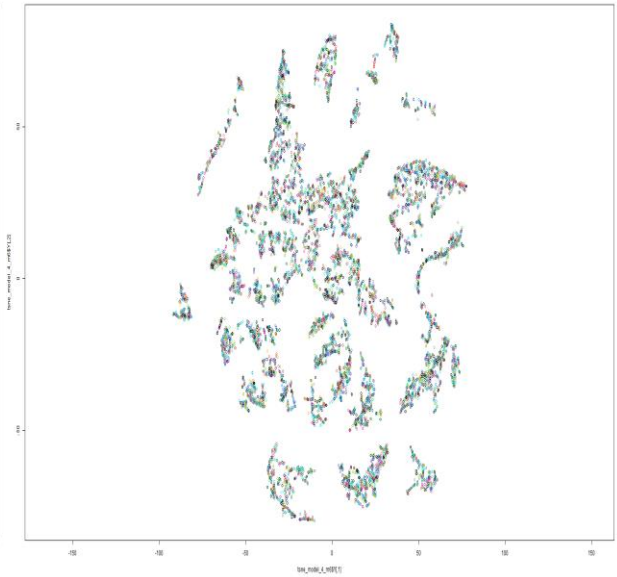


Figure 39 t-SNE on Model 6

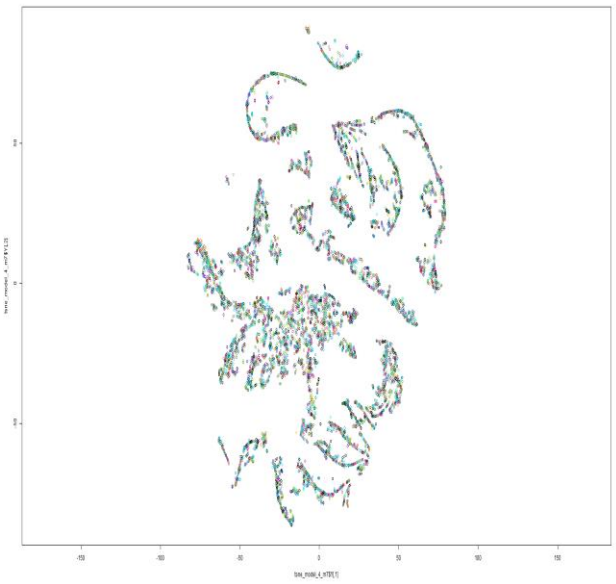


Figure 40 t-SNE on Model 7

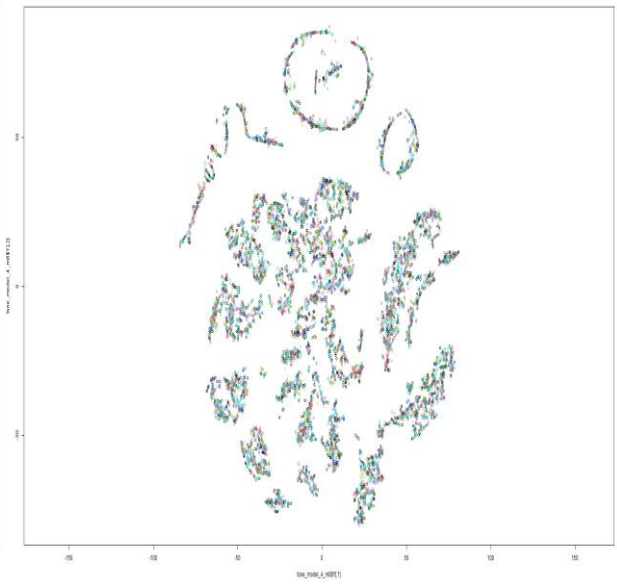


Figure 41 t-SNE on Model 8

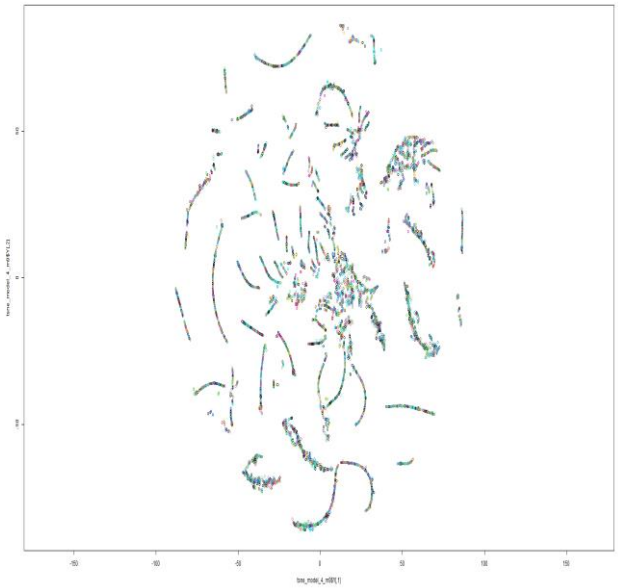


Figure 42 t-SNE on Model 9

Model 1:

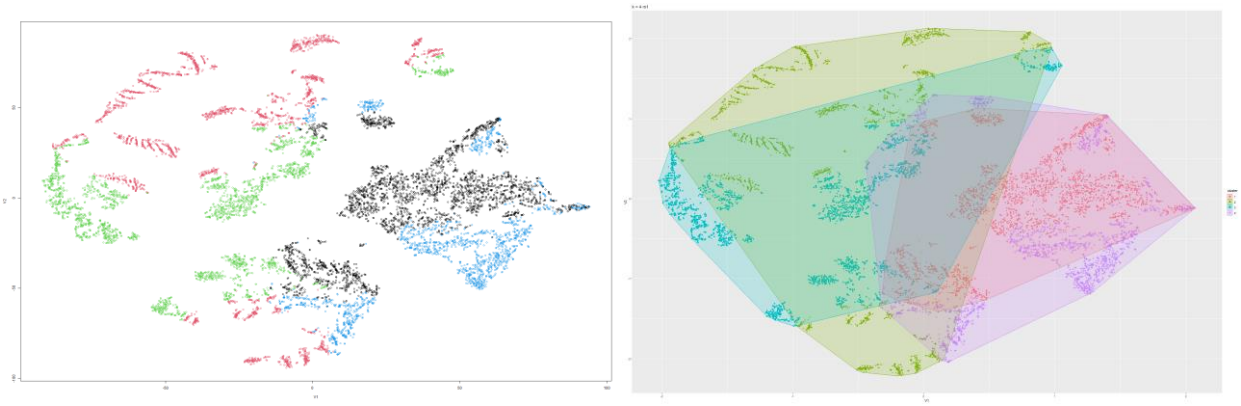


Figure 43&44 k-Means on t-SNE Model 1

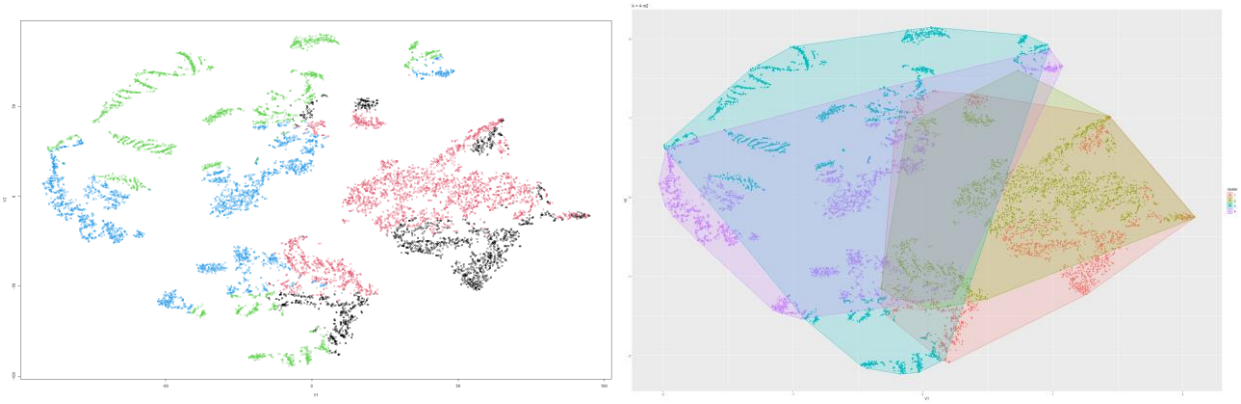


Figure 45 & 46 k-means t-SNE on Model 2

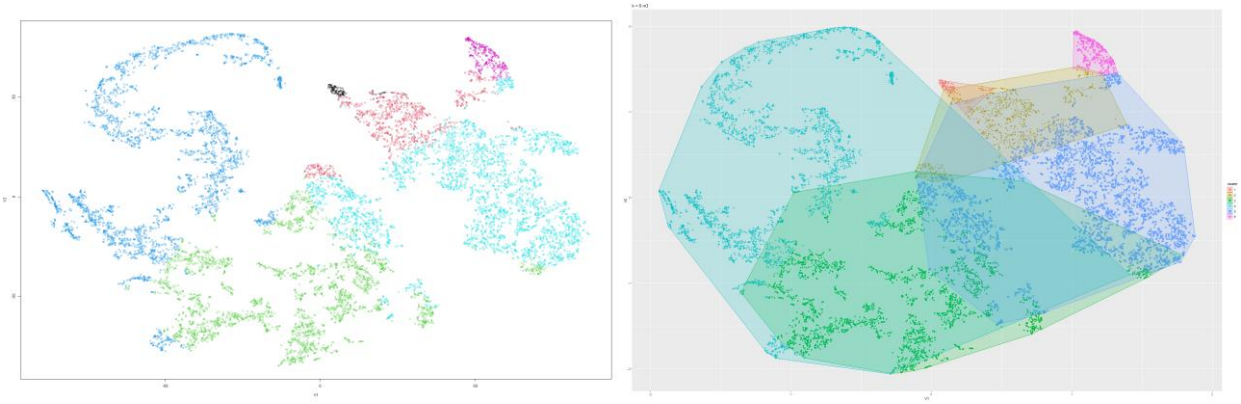


Figure 47 & 48 k-means t-SNE on Model 3

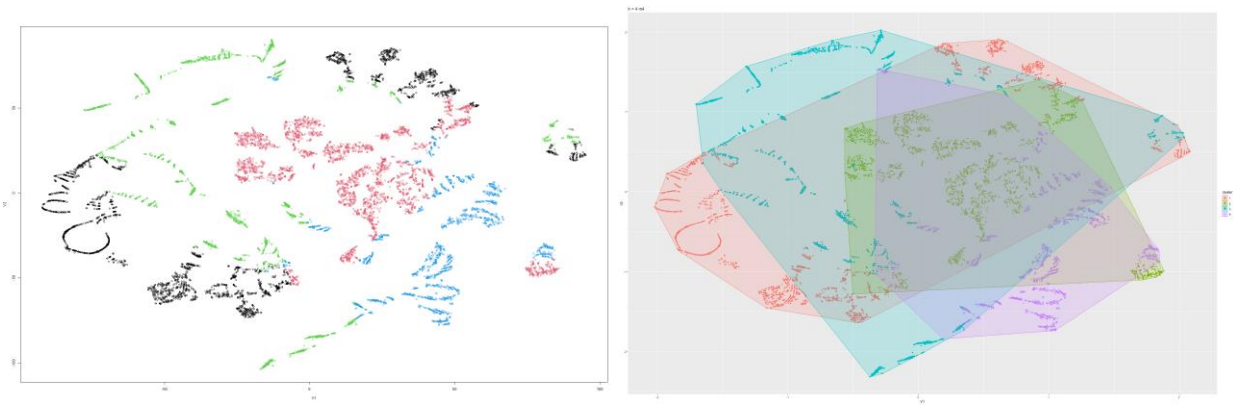


Figure 49 & 50 k-means t-SNE on Model 4

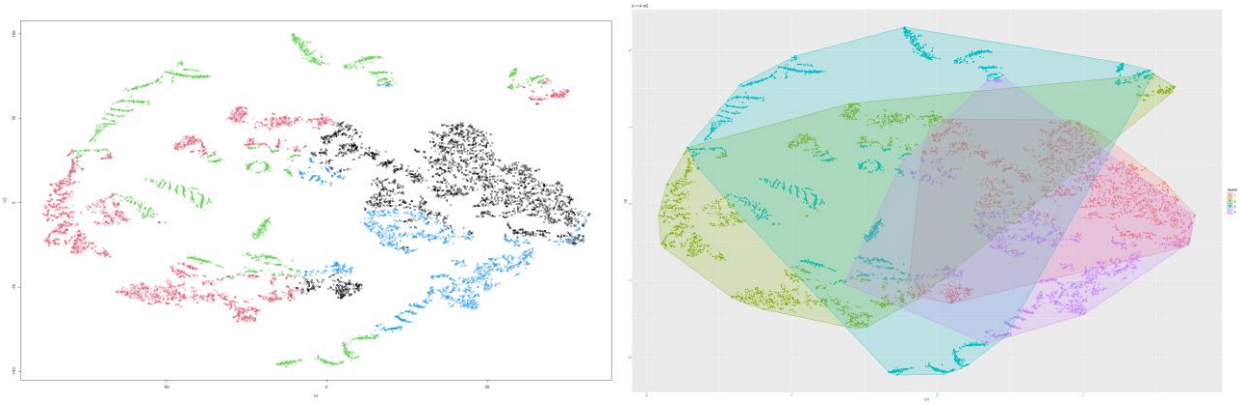


Figure 51 & 52 k-means t-SNE on Model 5

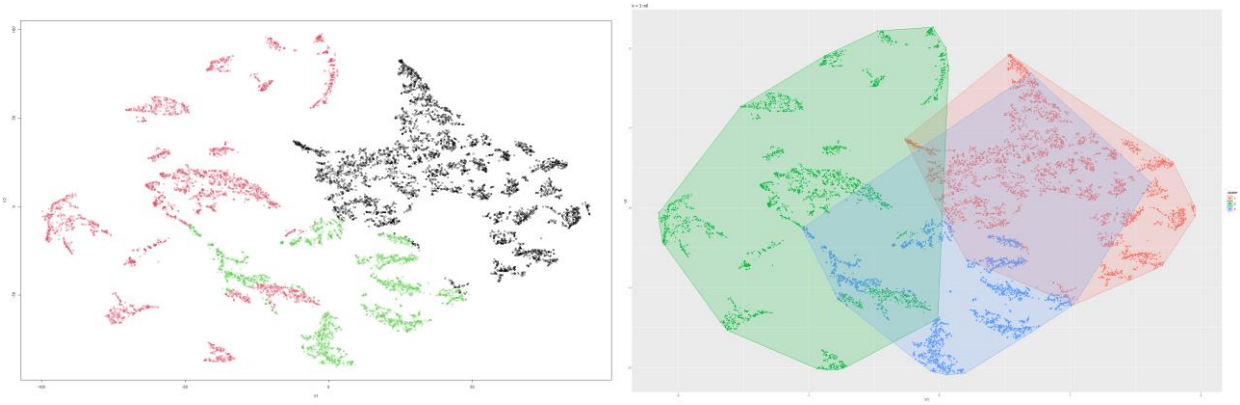


Figure 53 & 54 k-means t-SNE on Model 6

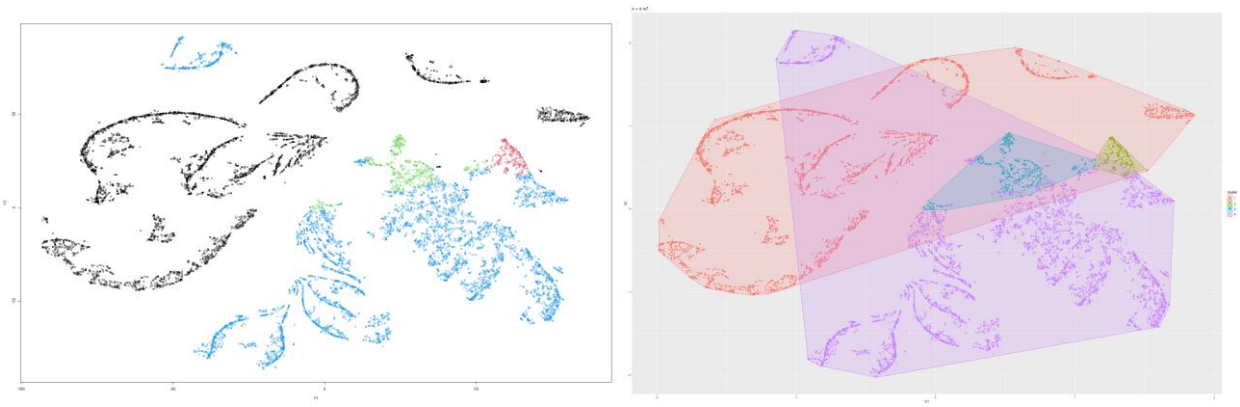


Figure 55 & 56 k-means t-SNE on Model 7

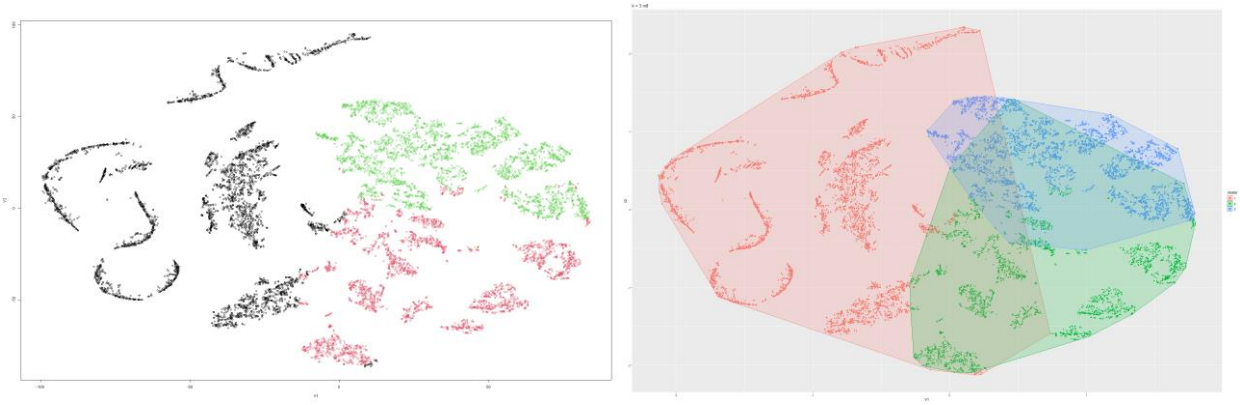


Figure 57 & 58 k-means t-SNE on Model 8

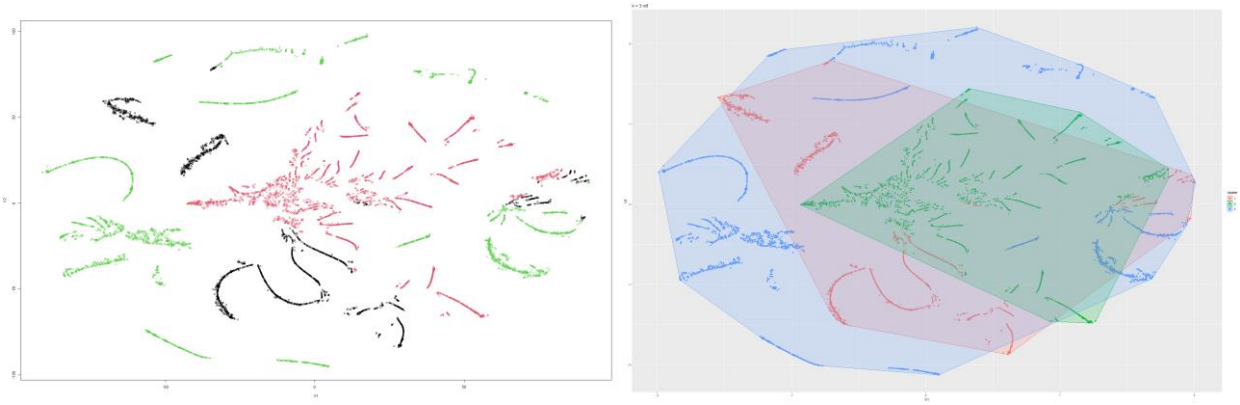


Figure 59 & 60 k-means t-SNE on Model 2

Model 2:

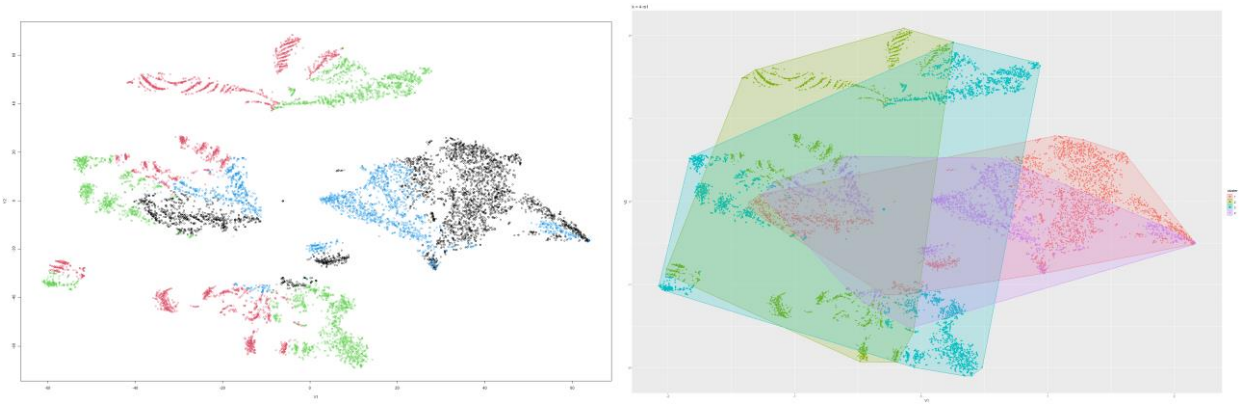


Figure 61 & 62 k-means t-SNE on Model 1

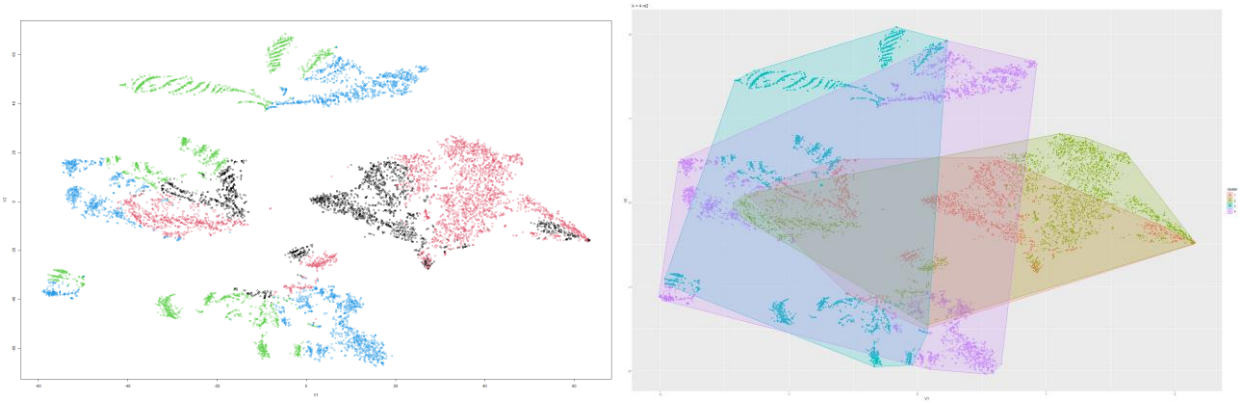


Figure 63 & 64 k-means t-SNE on Model 2

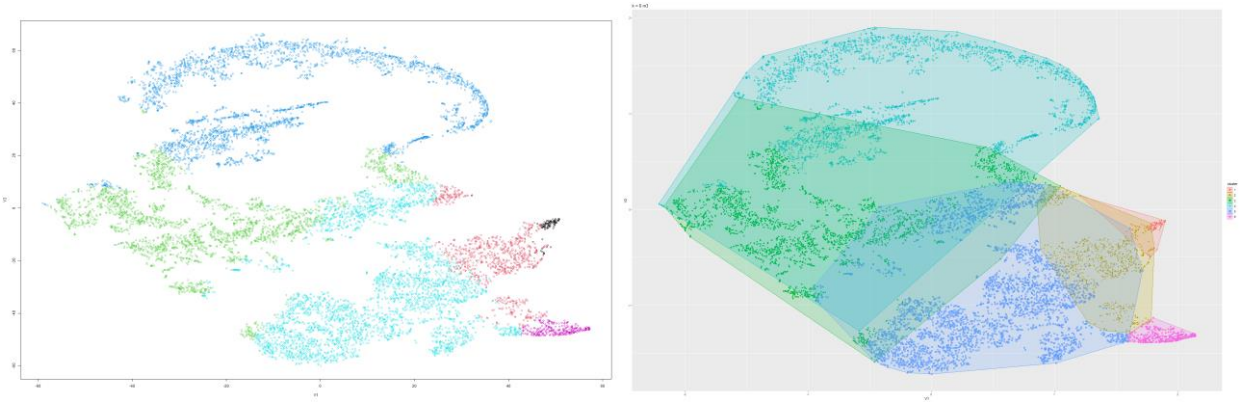


Figure 65 & 66 k-means t-SNE on Model 3

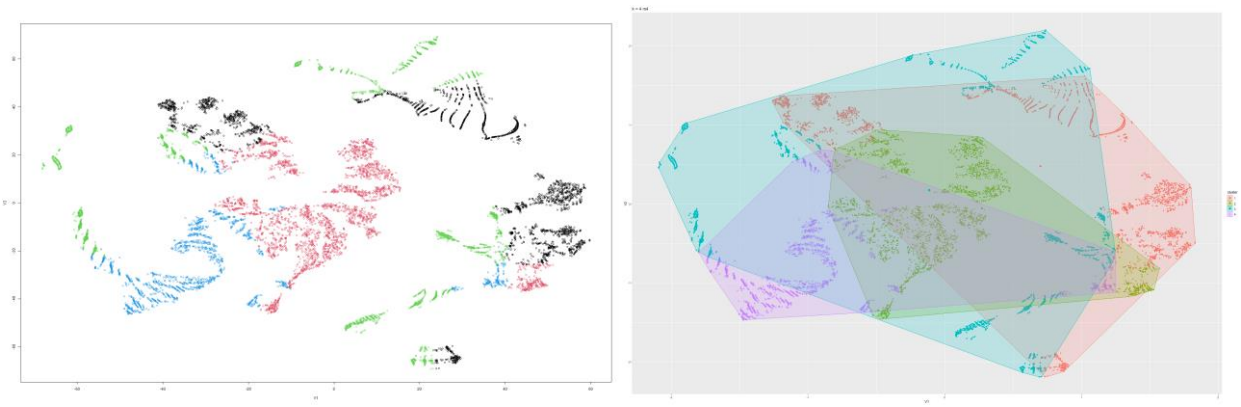


Figure 67 & 68 k-means t-SNE on Model 4

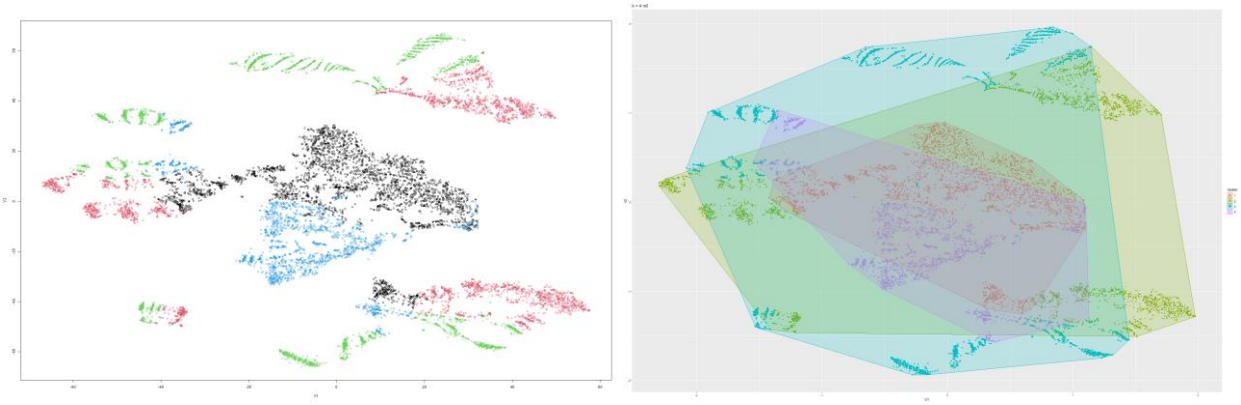


Figure 69 & 70 k-means t-SNE on Model 5

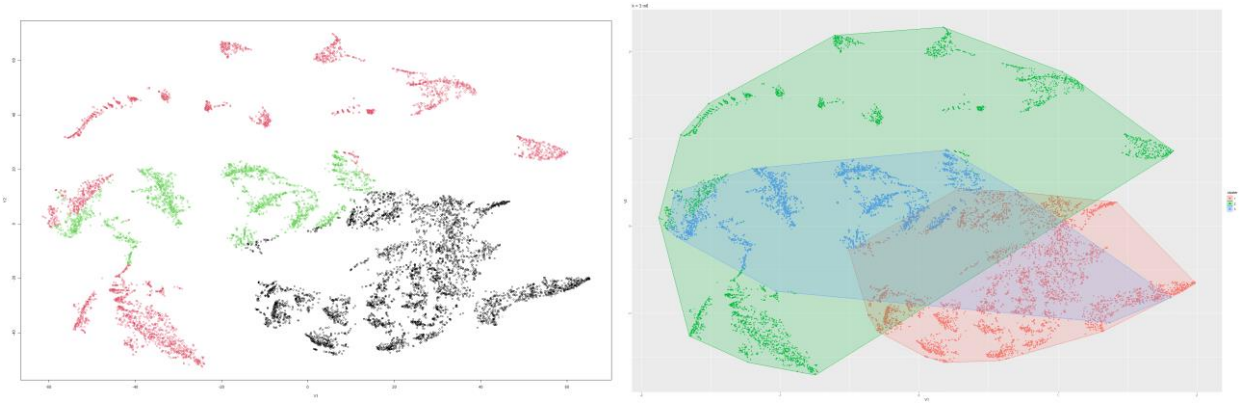


Figure 71 & 72 k-means t-SNE on Model 6

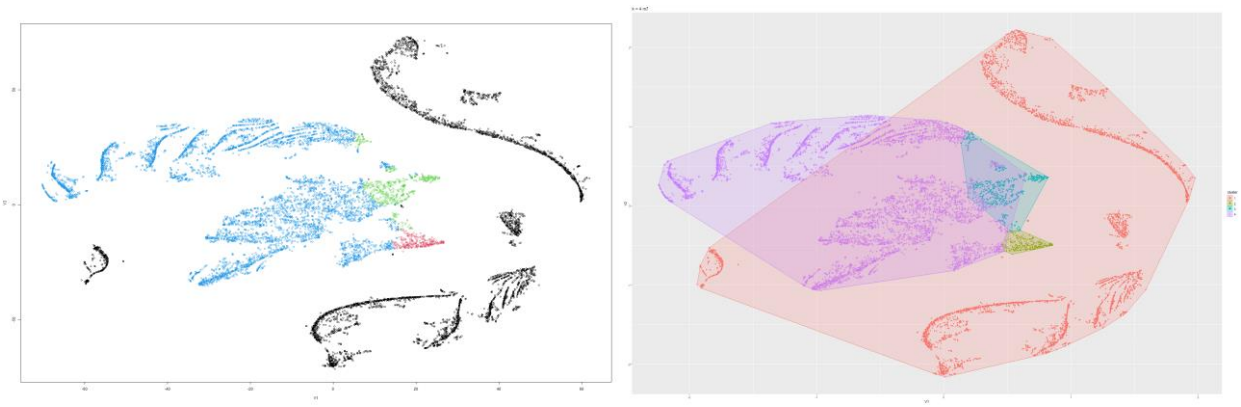


Figure 73 & 74 k-means t-SNE on Model 7

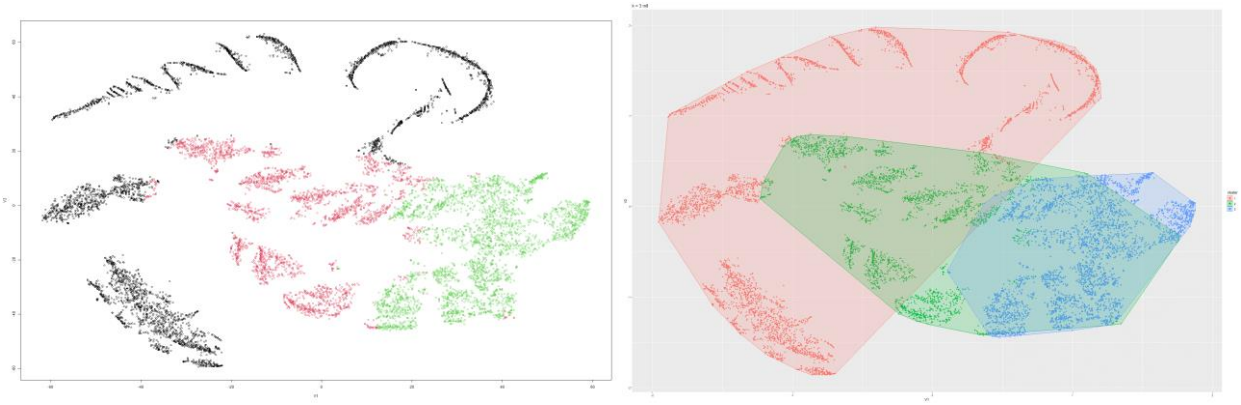


Figure 75 & 76 k-means t-SNE on Model 8

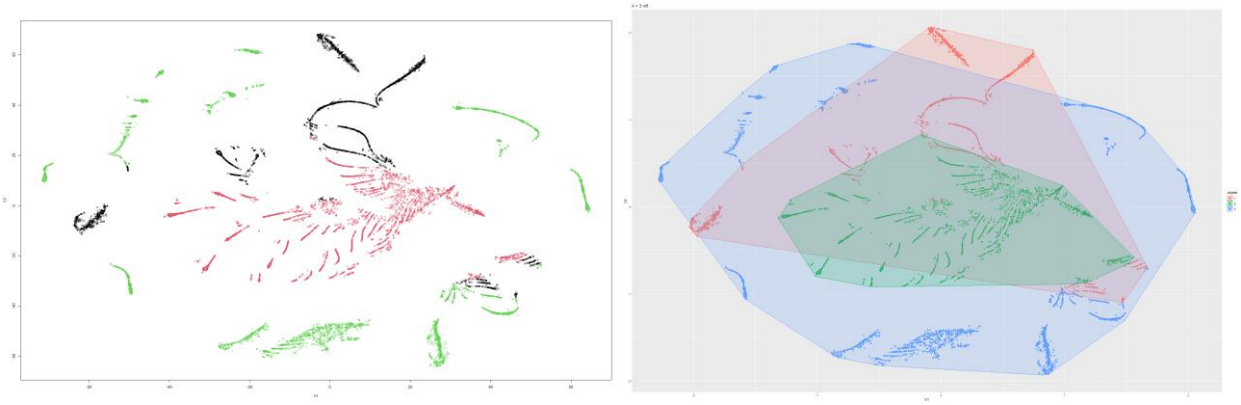


Figure 77 & 78 k-means t-SNE on Model 9

Model 3:

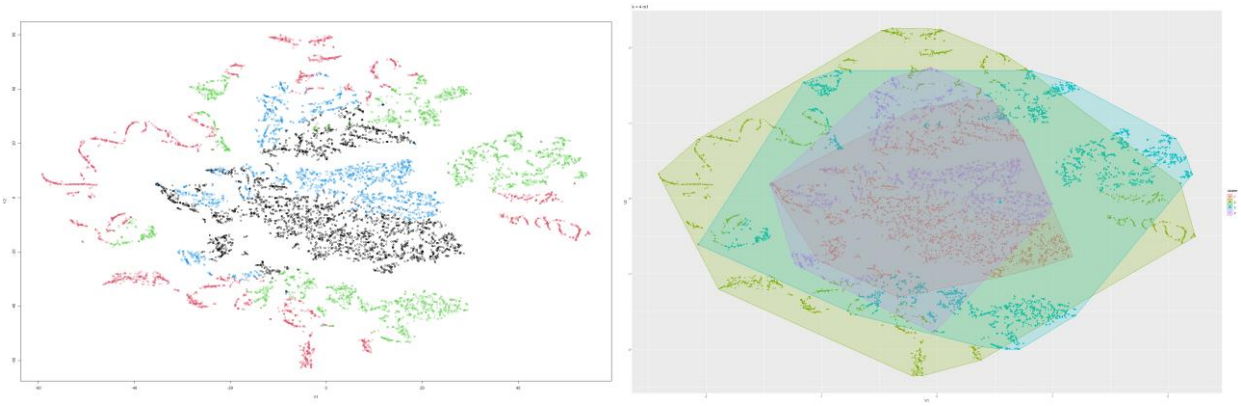


Figure 79 & 80 k-means t-SNE on Model 1

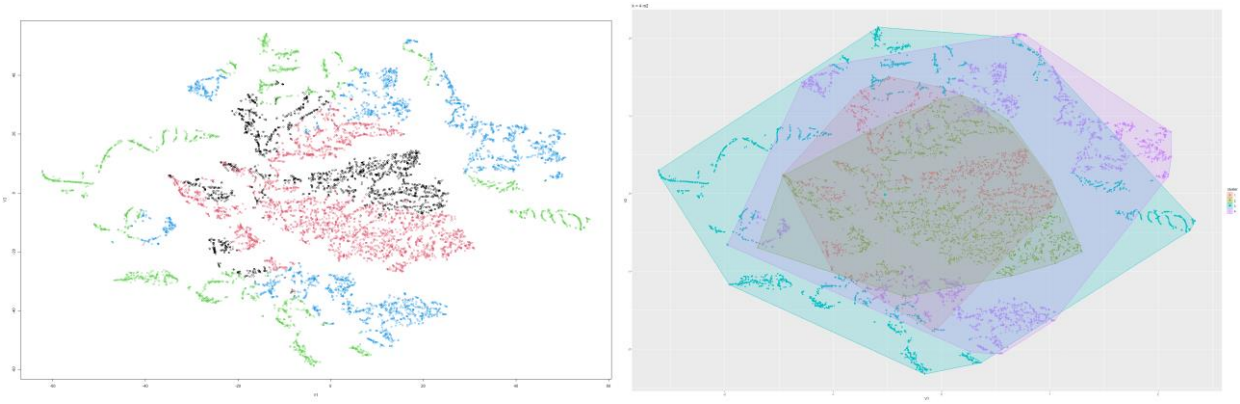


Figure 81 & 82 k-means t-SNE on Model 2

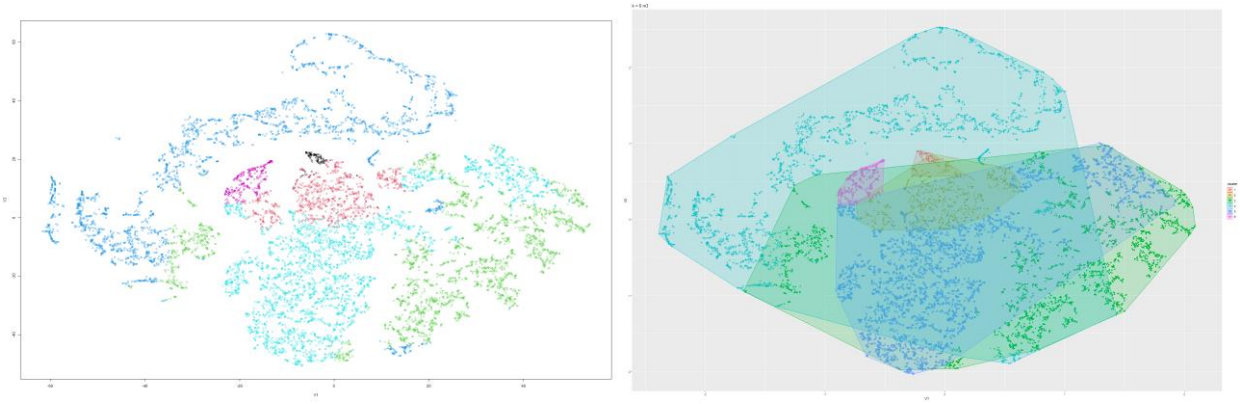


Figure 83 & 84 k-means t-SNE on Model 3

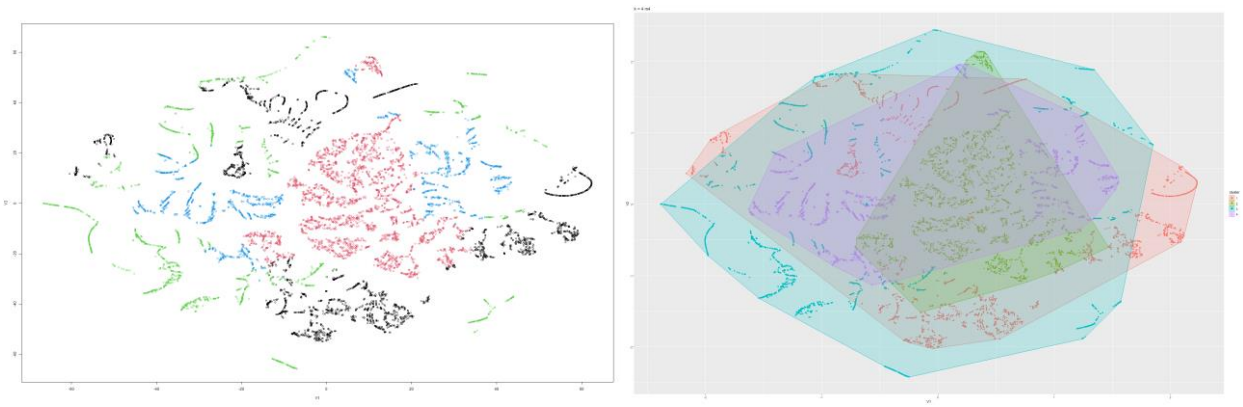


Figure 85 & 86 k-means t-SNE on Model 4

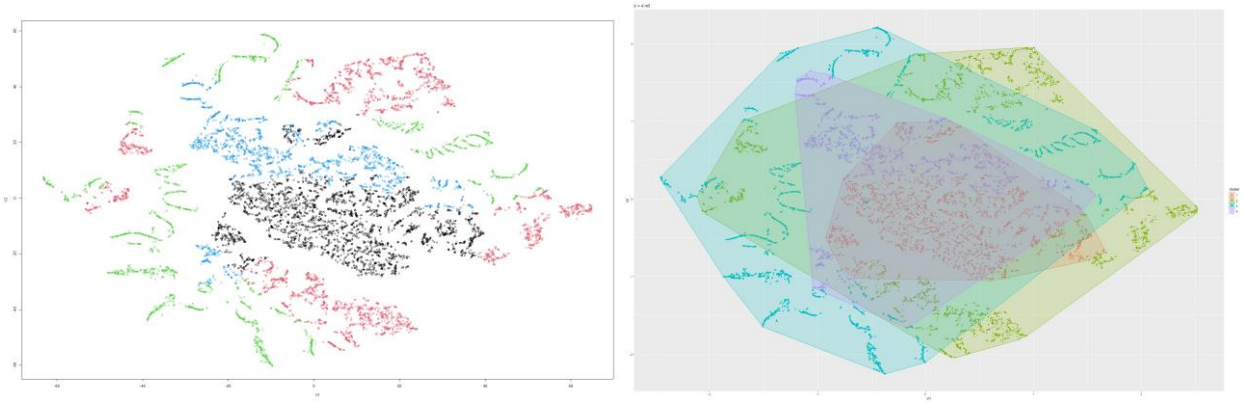


Figure 87 & 88 k-means t-SNE on Model 5

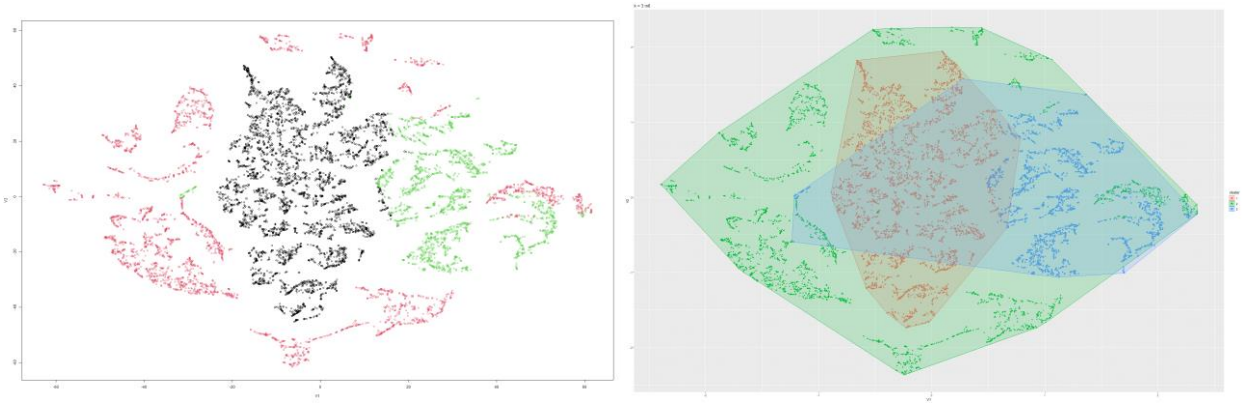


Figure 89 & 90 k-means t-SNE on Model 6

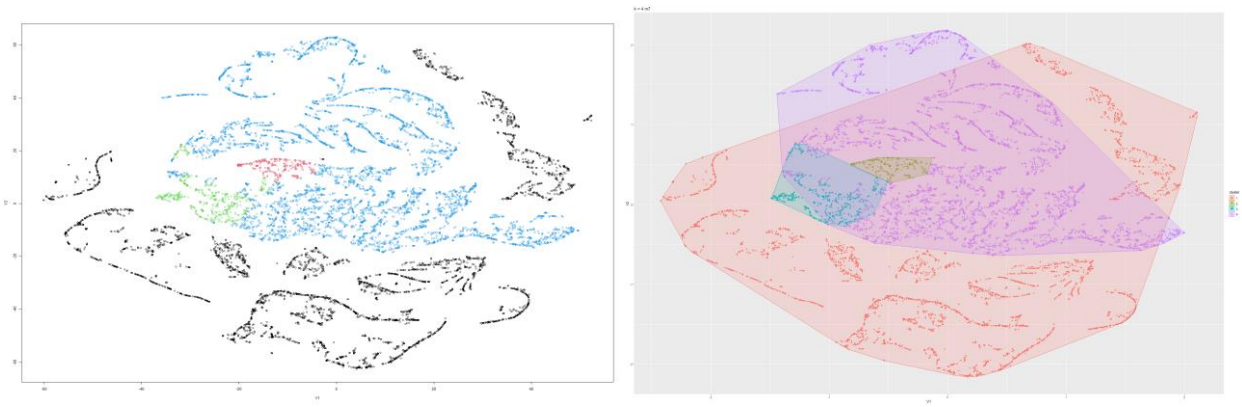


Figure 91 & 92 k-means t-SNE on Model 7

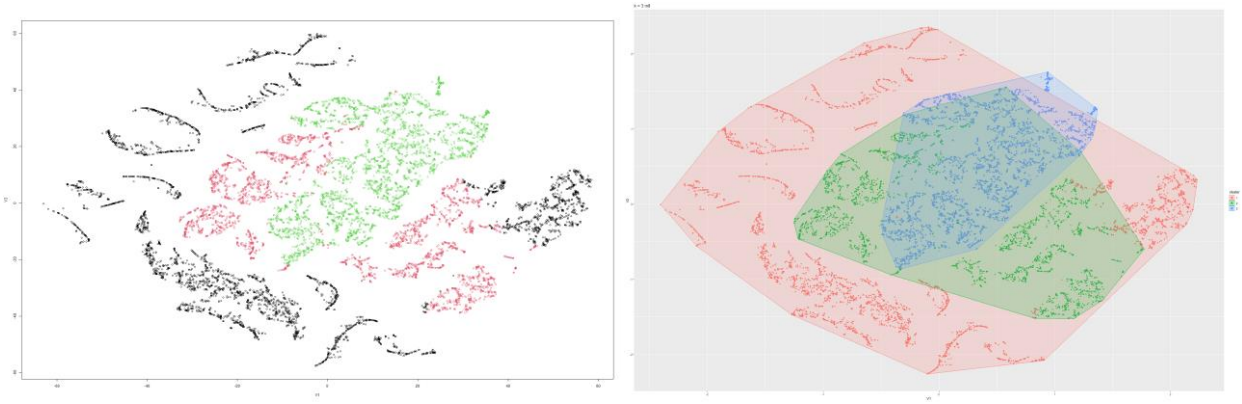


Figure 93 & 94 k-means t-SNE on Model 8

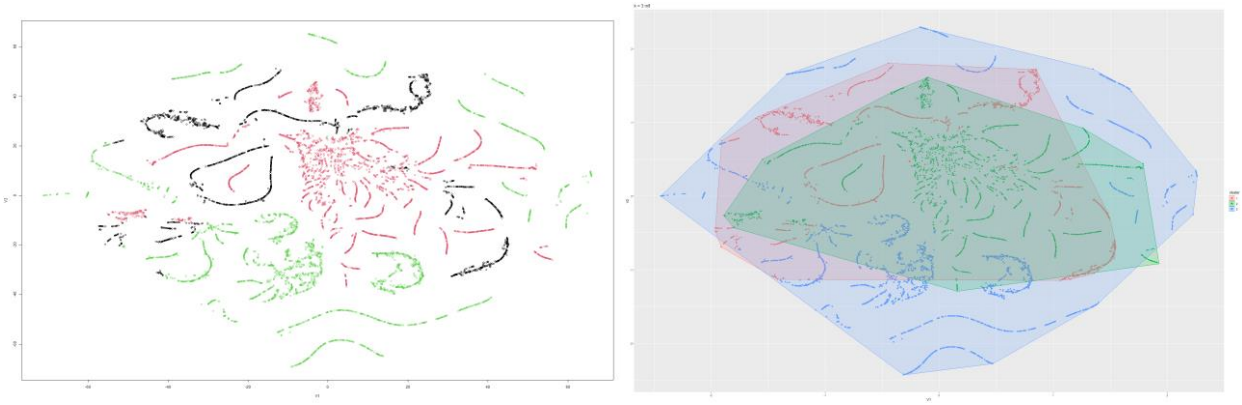


Figure 95 & 96 k-means t-SNE on Model 9

Model 4:

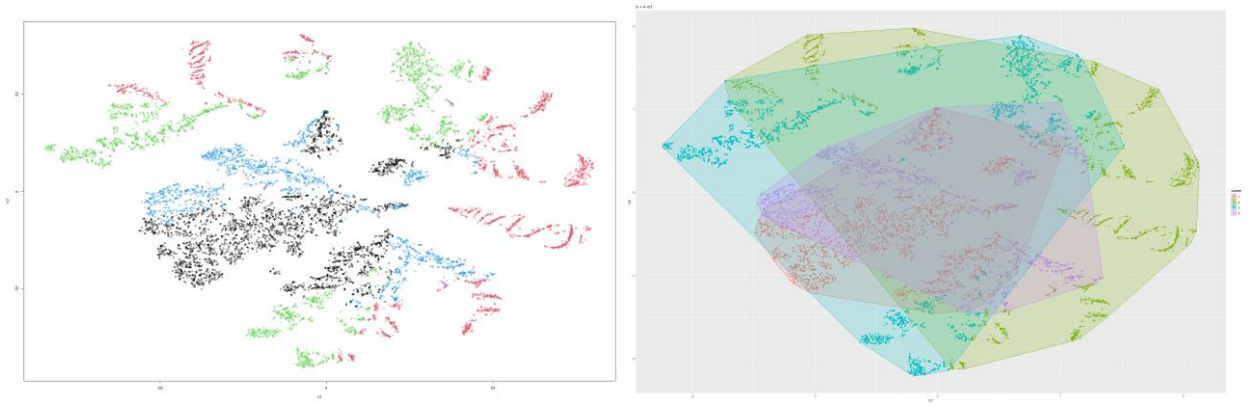


Figure 97 & 98 k-means t-SNE on Model 1

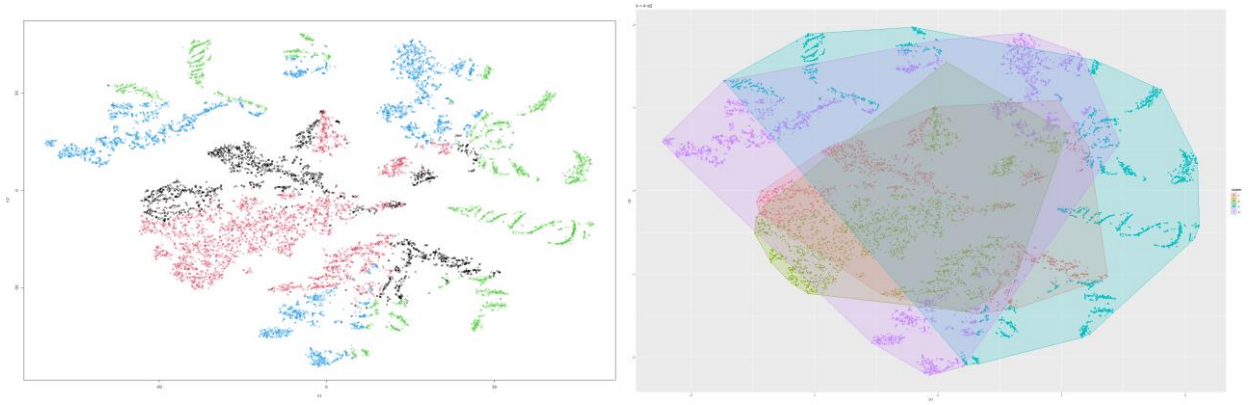


Figure 99 & 100 k-means t-SNE on Model 2

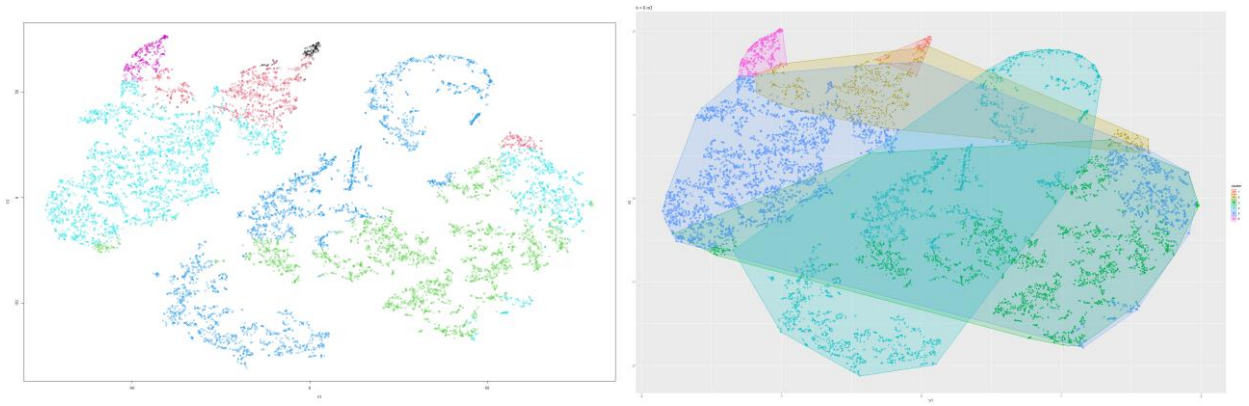


Figure 101 & 102 k-means t-SNE on Model 3

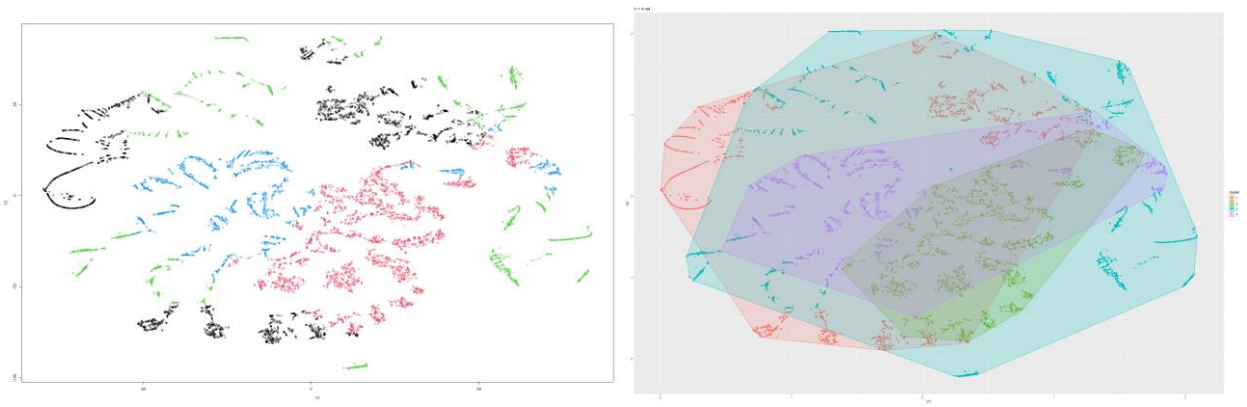


Figure 103 & 104 k-means t-SNE on Model 4

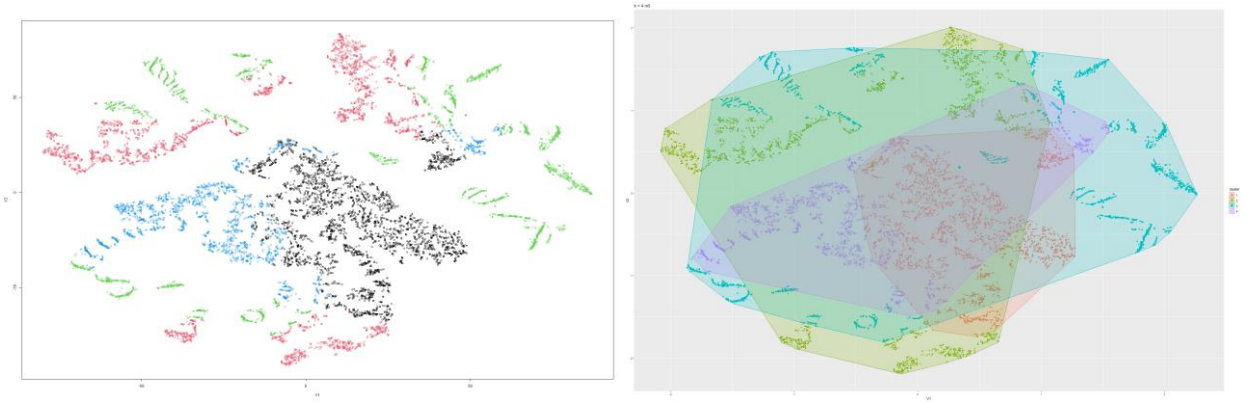


Figure 105 & 106 k-means t-SNE on Model 5

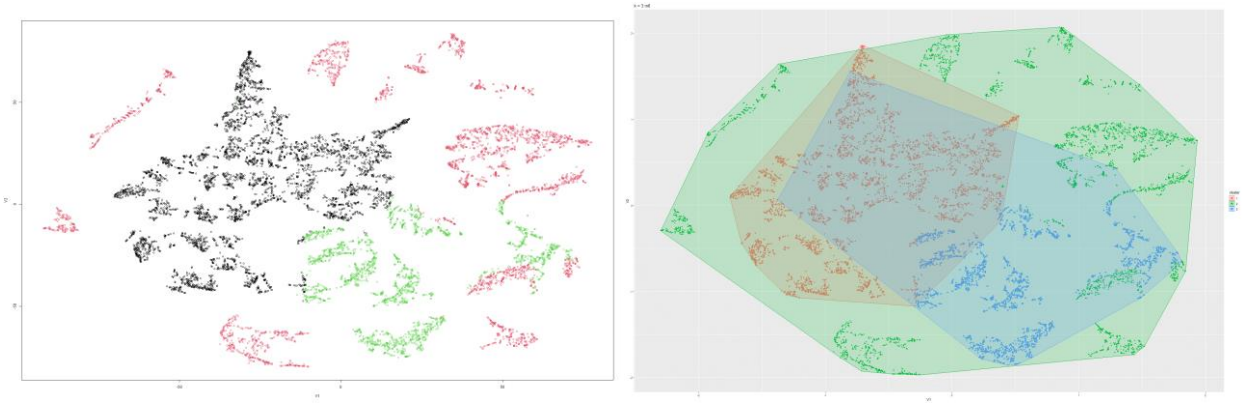


Figure 107 & 108 k-means t-SNE on Model 6

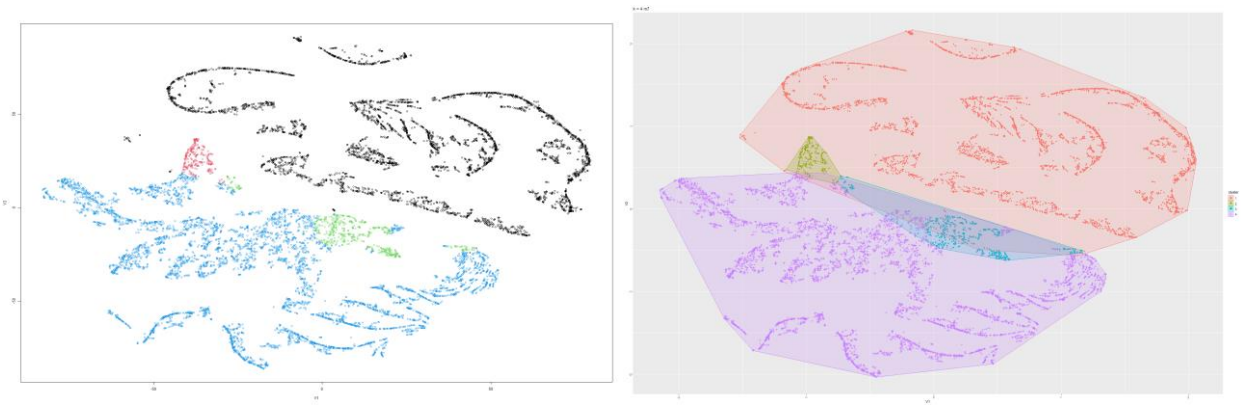


Figure 109 & 110 k-means t-SNE on Model 7

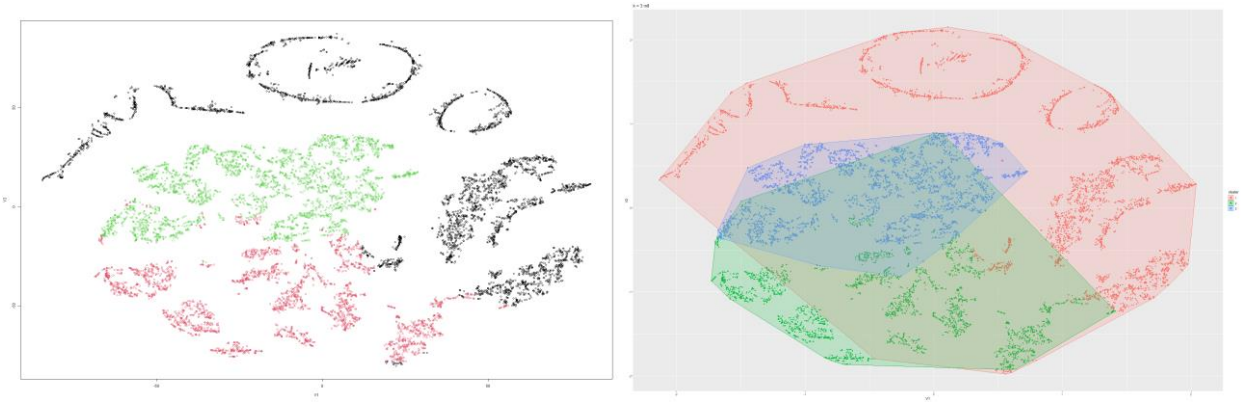


Figure 111 & 112 k-means t-SNE on Model 8

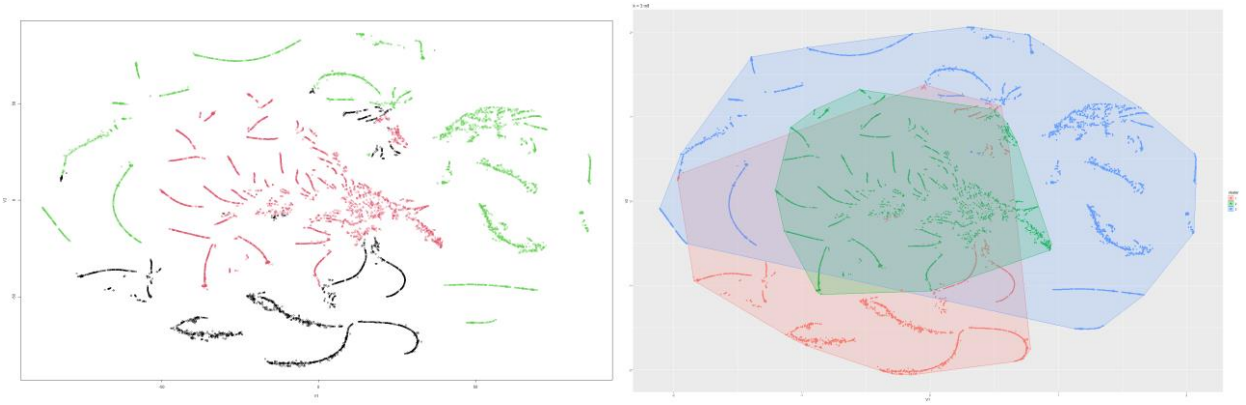


Figure 113 & 114 k-means t-SNE on Model 9

6.1.4 K-means on PCA

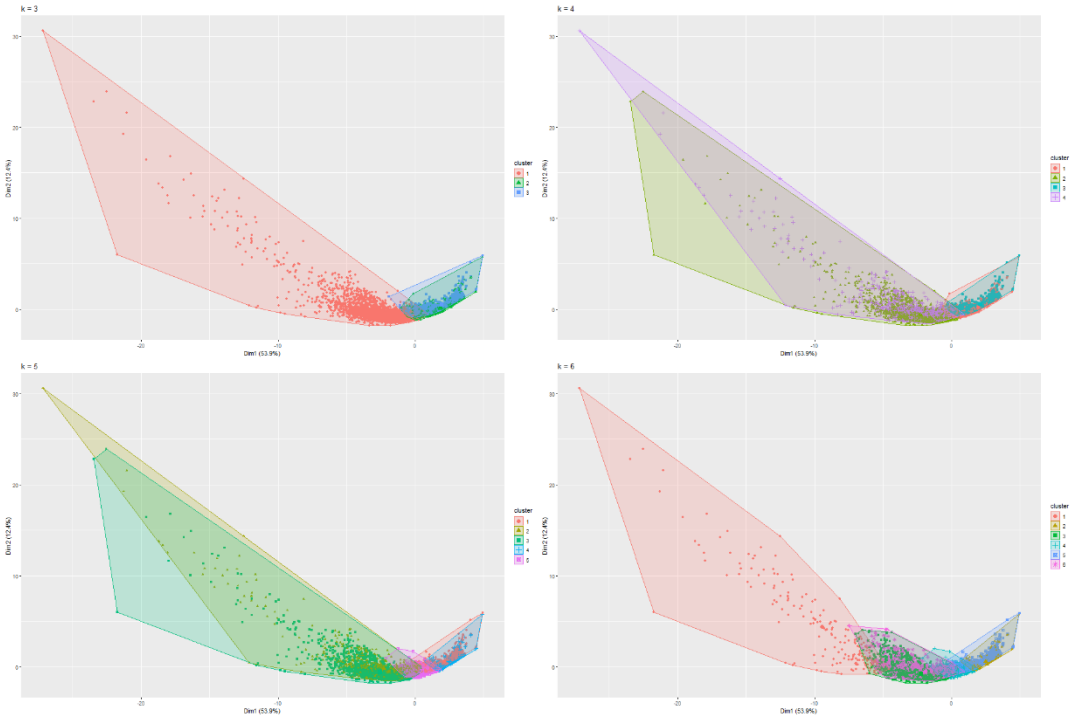


Figure 115 k-means PCA on Model 1

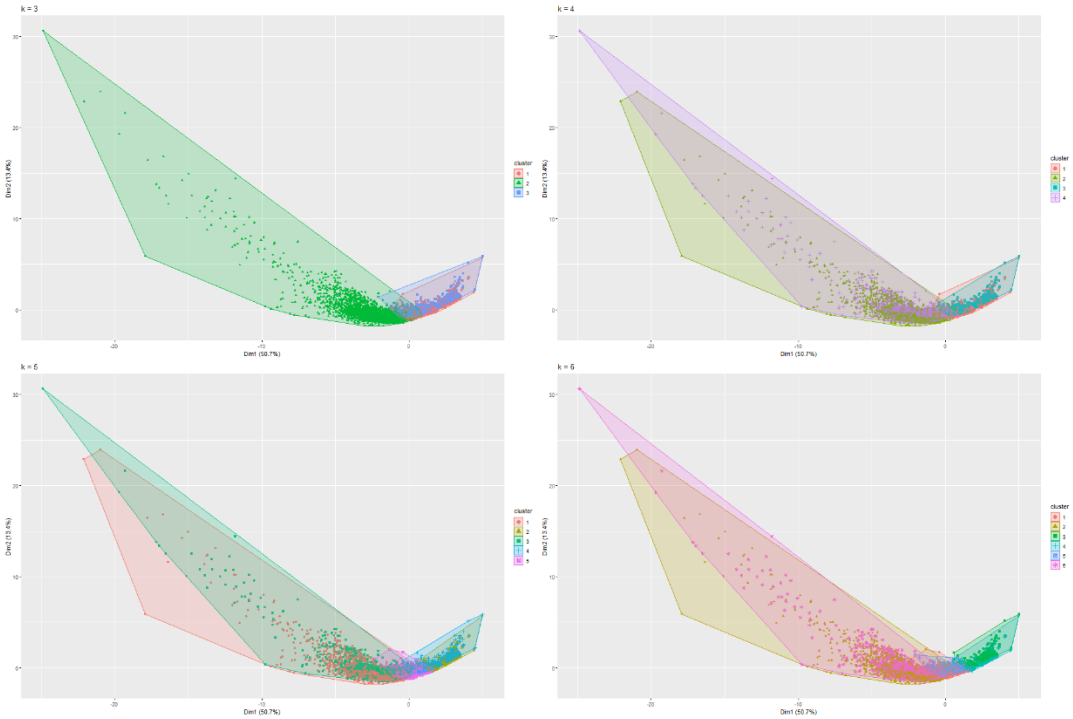


Figure 116 k-means PCA on Model 2

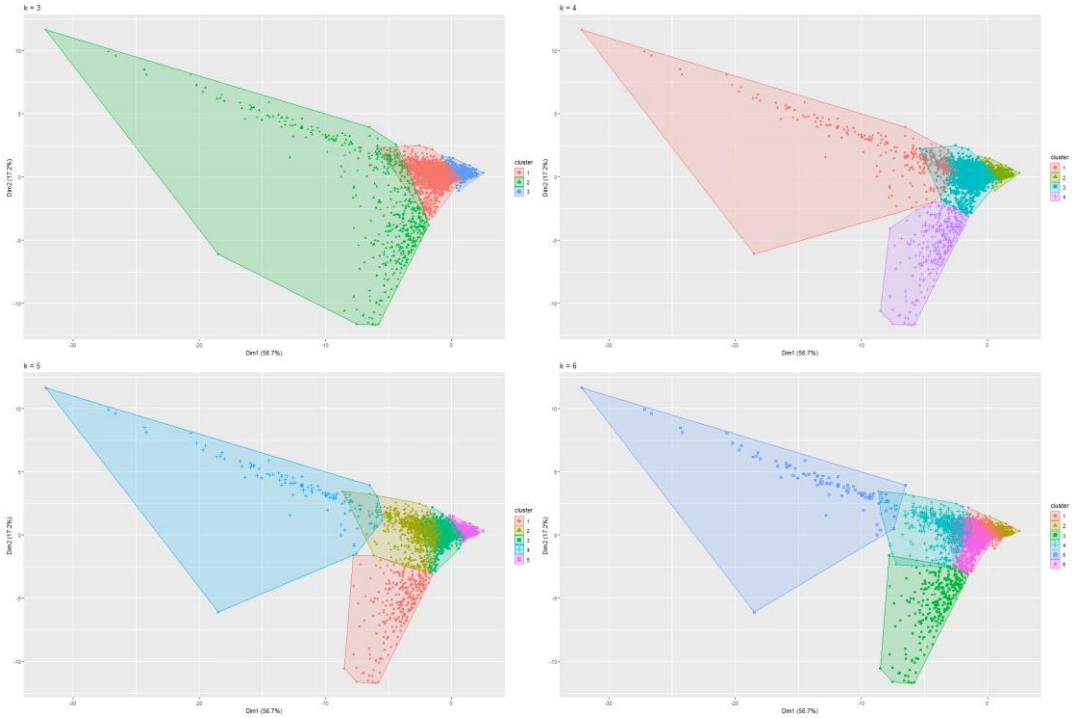


Figure 117 k-means PCA on Model 3

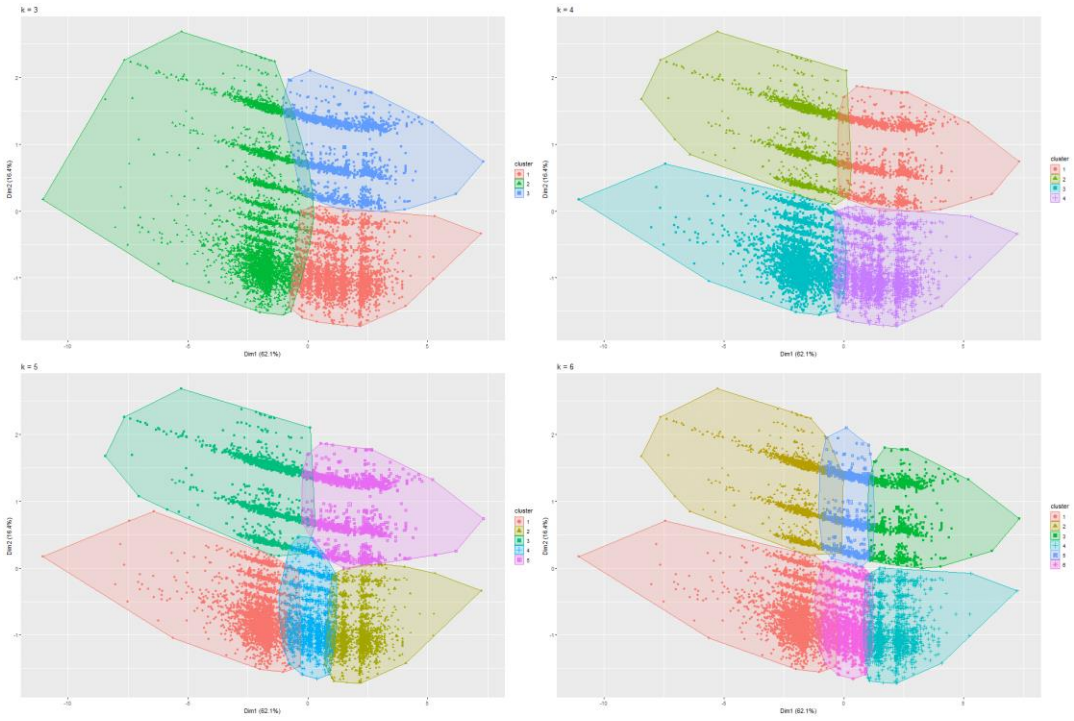


Figure 118 k-means PCA on Model 4

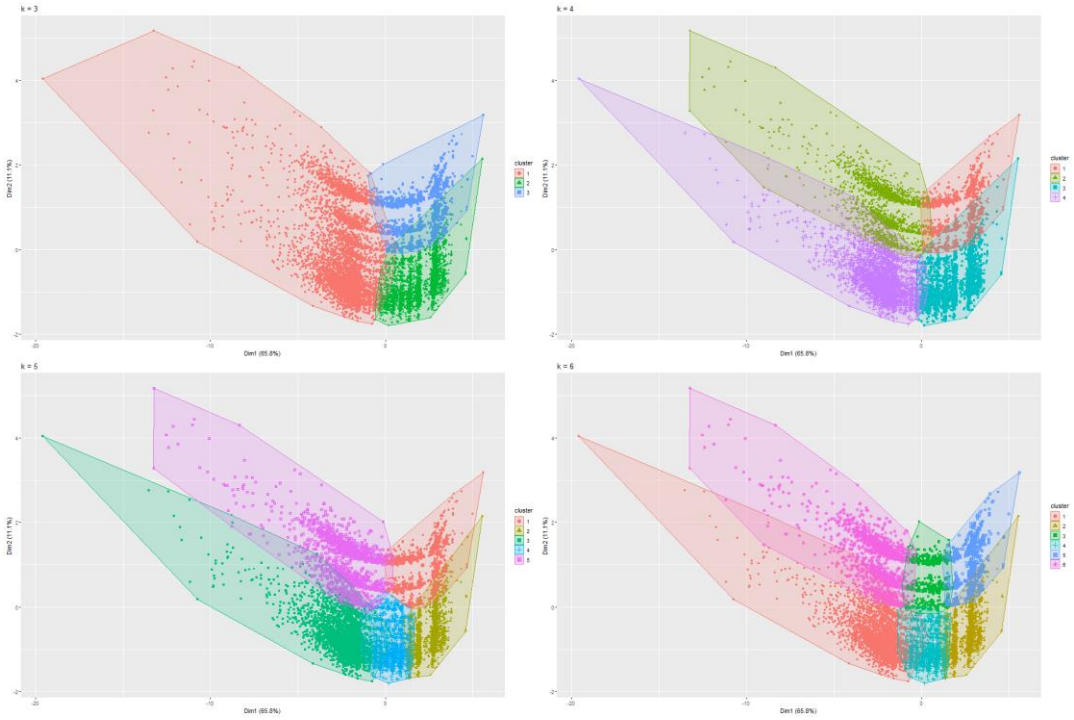


Figure 119 k-means PCA on Model 5

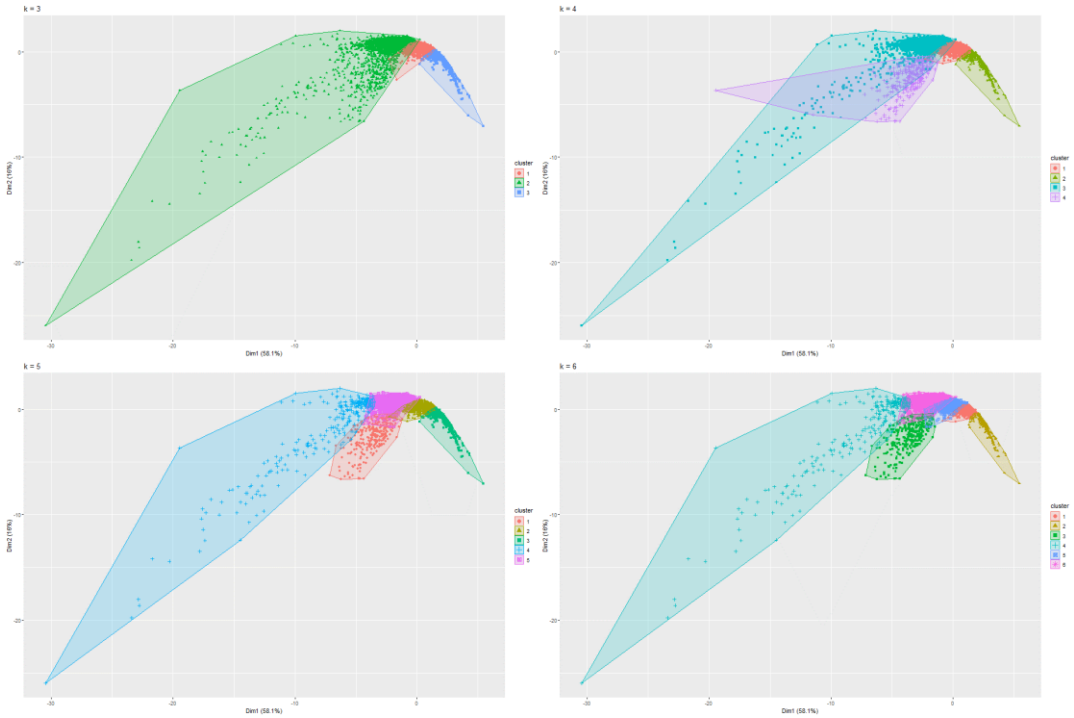


Figure 120 k-means PCA on Model 6

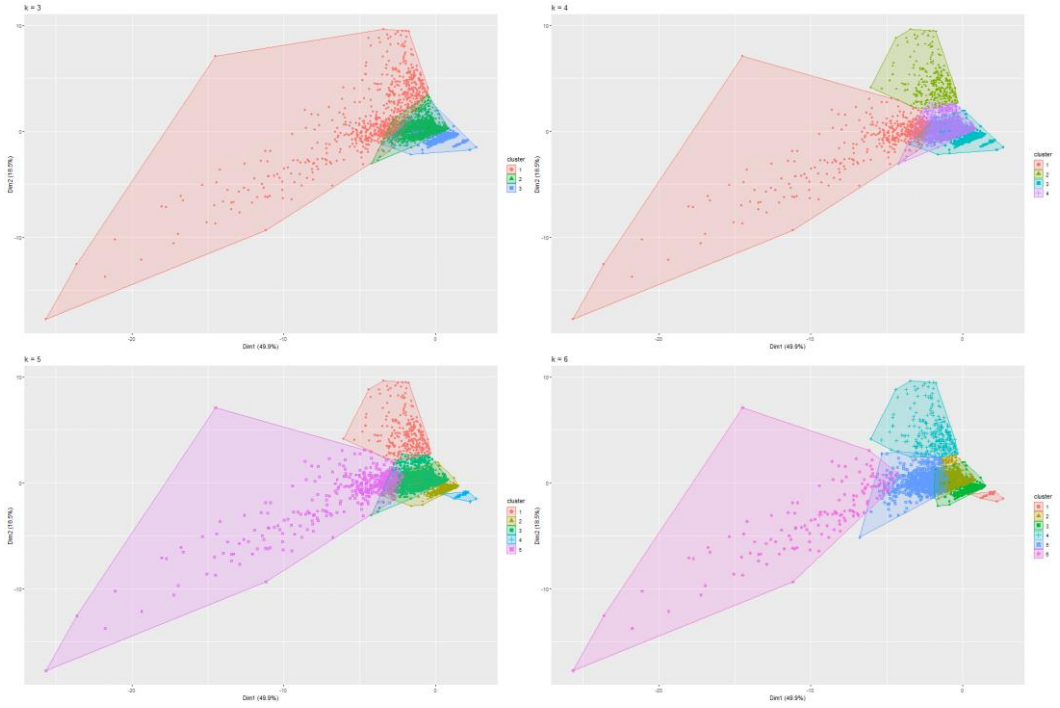


Figure 121 k-means PCA on Model 7

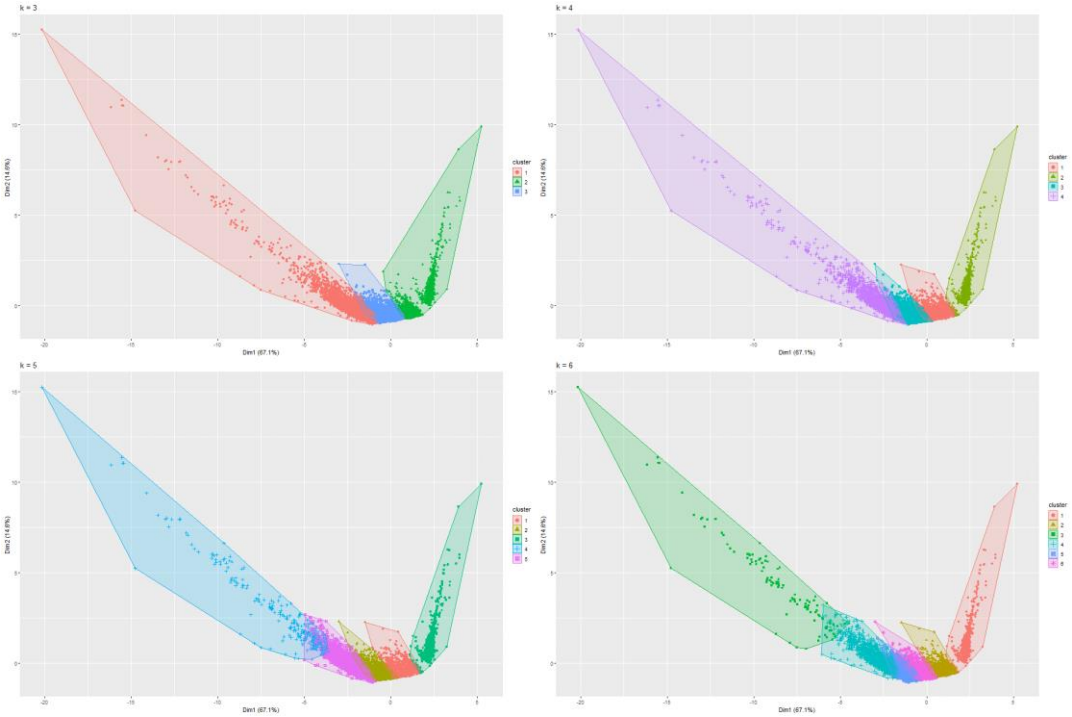


Figure 122 k-means PCA on Model 8

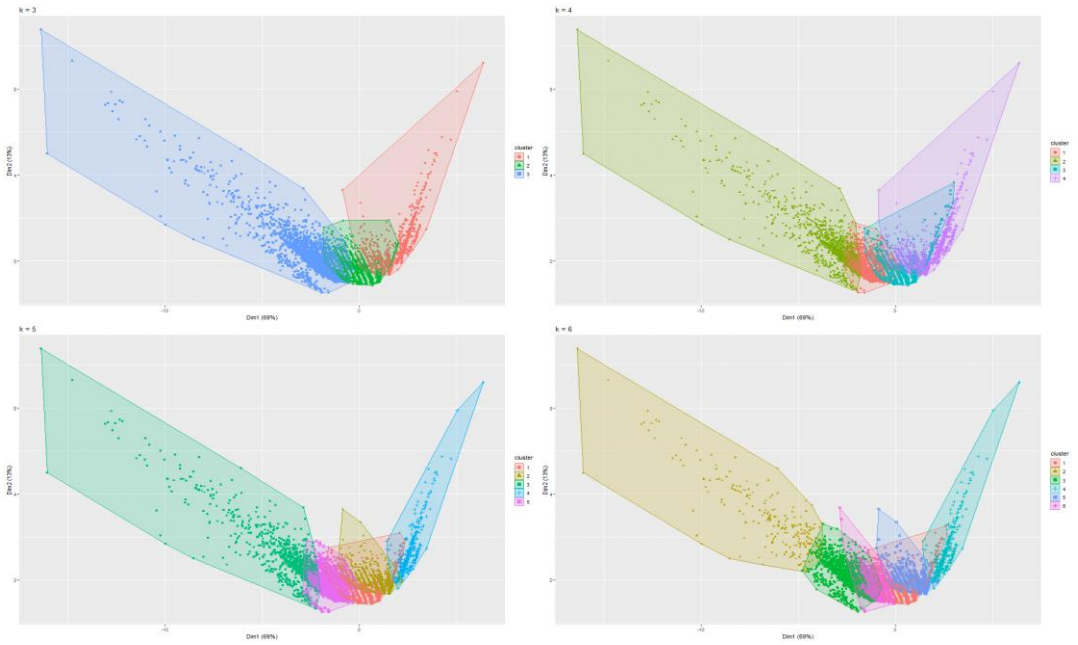


Figure 122 k-means PCA on Model 8

6.1.5 UMAP and K-means on UMAP

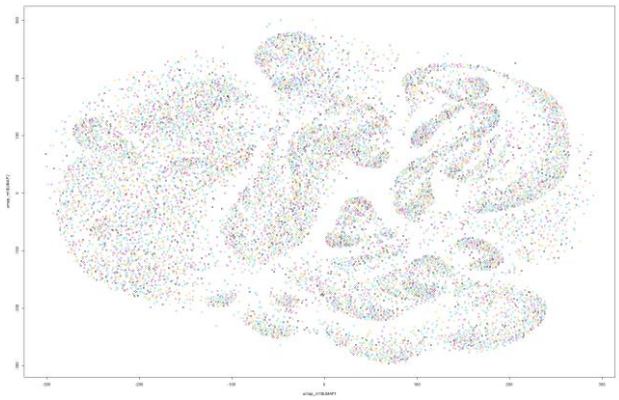


Figure 123 UMAP on Model 1

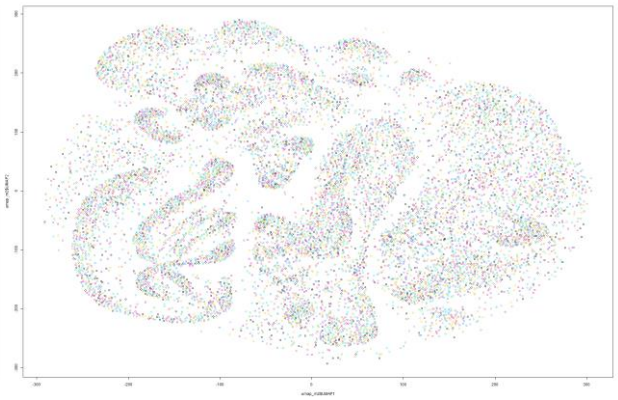


Figure 124 UMAP on Model 2

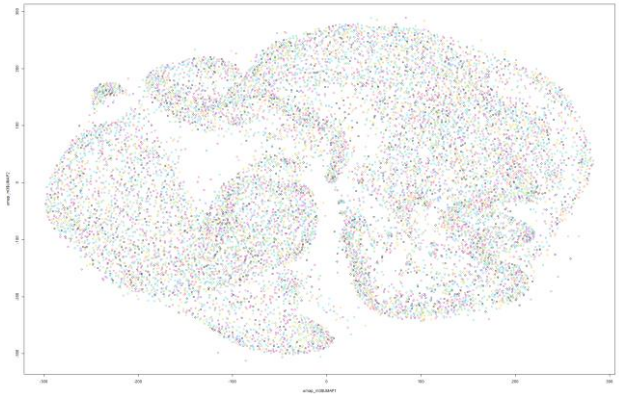


Figure 125 UMAP on Model 3

Figure 126 UMAP on Model 4

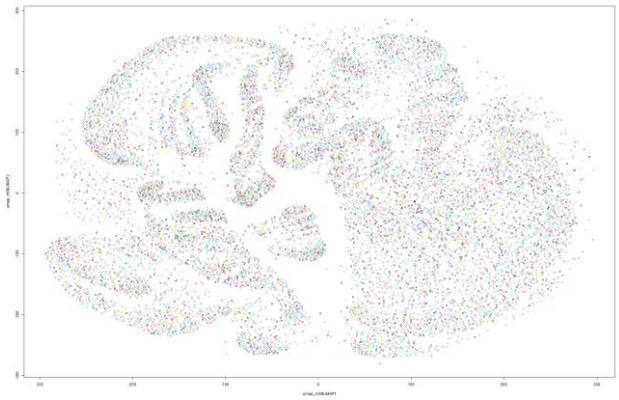


Figure 127 UMAP on Model 5

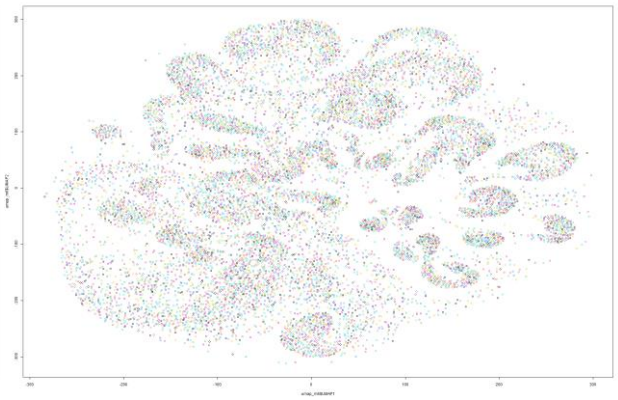


Figure 128 UMAP on Model 6

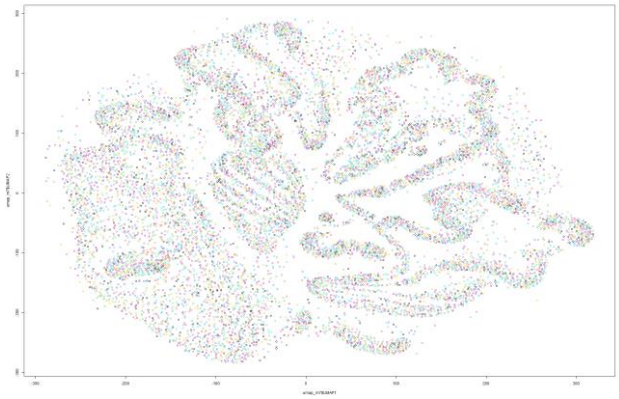


Figure 129 UMAP on Model 7

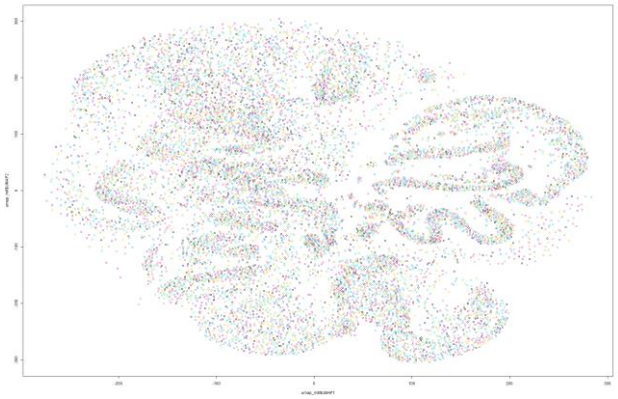


Figure 130 UMAP on Model 8

Figure 131 UMAP on Model 9

K-means on UMAP

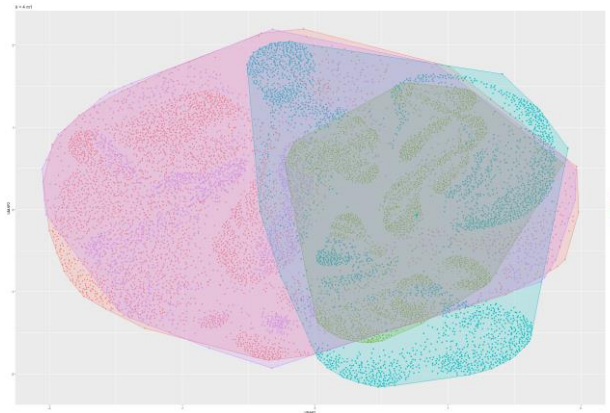


Figure 132 k-Means UMAP on Model 1

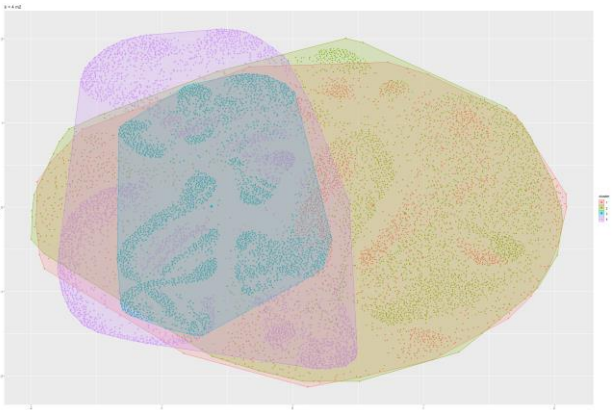


Figure 133 k-Means UMAP on Model 2

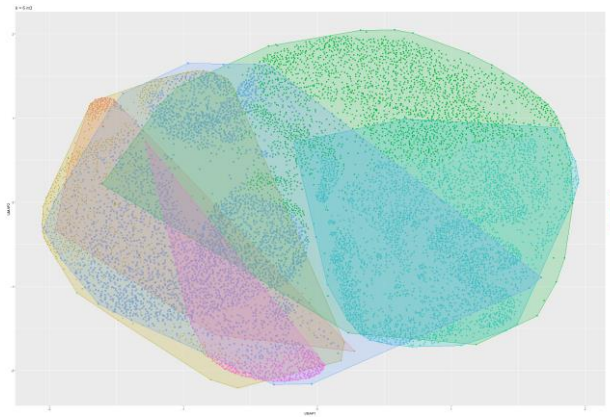


Figure 134 k-Means UMAP on Model 3

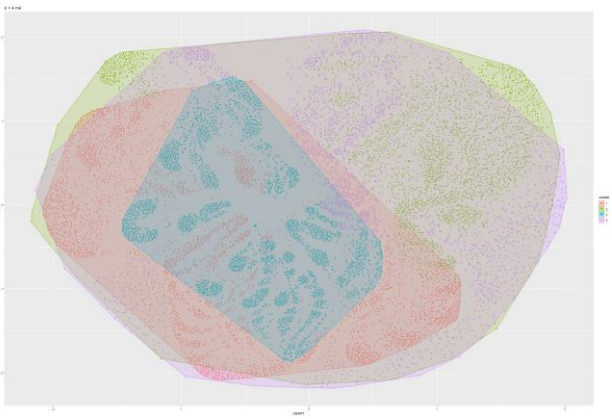


Figure 135 k-Means UMAP on Model 4

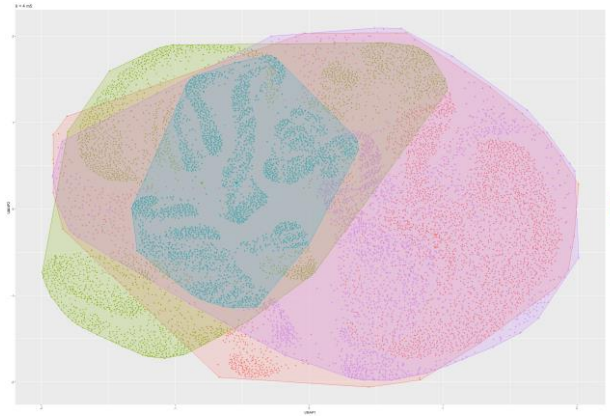


Figure 136 k-Means UMAP on Model 5

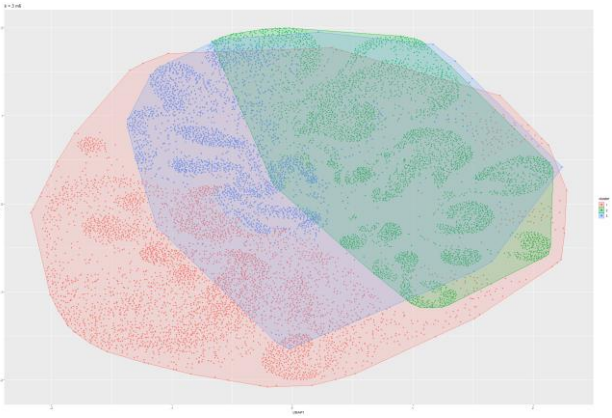


Figure 137 k-Means UMAP on Model 6

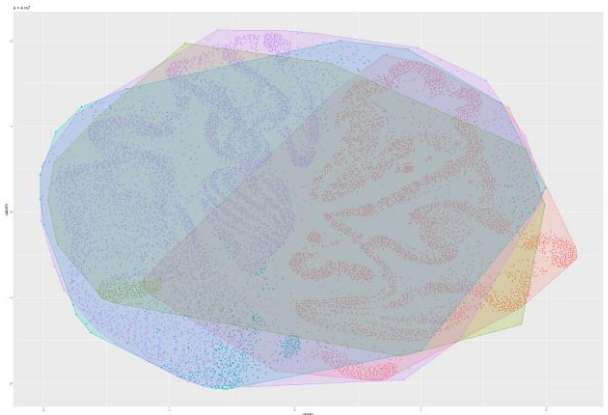


Figure 138 k-Means UMAP on Model 7

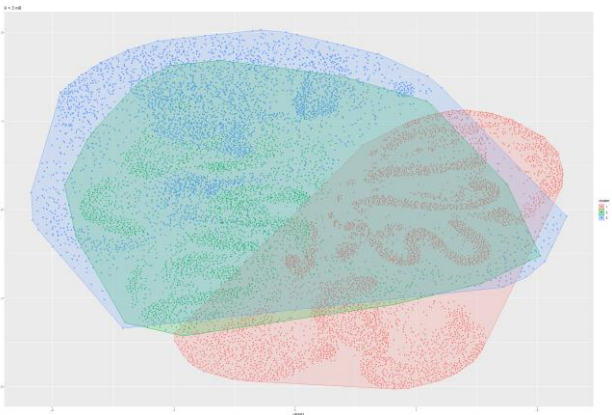


Figure 139 k-Means UMAP on Model 8

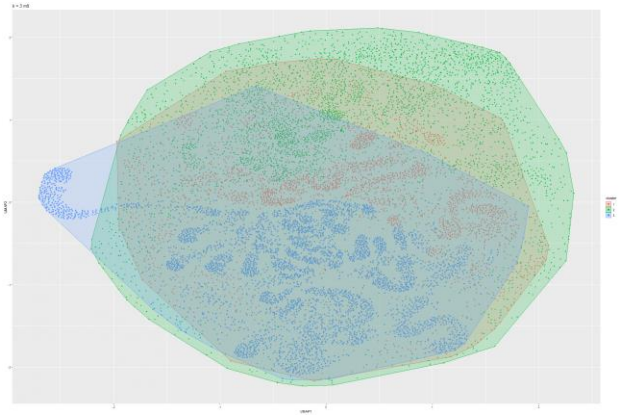


Figure 140 k-Means UMAP on Model 9

6.1.6 Sammon's mapping

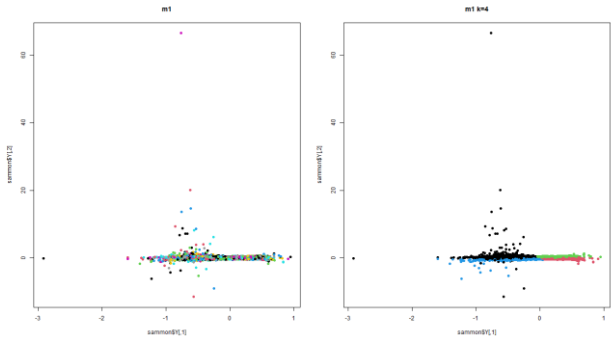


Figure 141 Sammon's mapping on Model 1

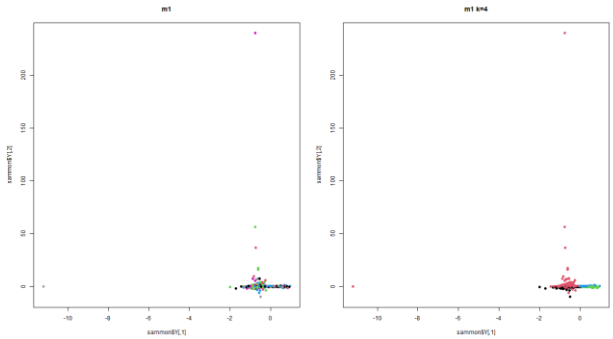


Figure 142 Sammon's mapping on Model 2

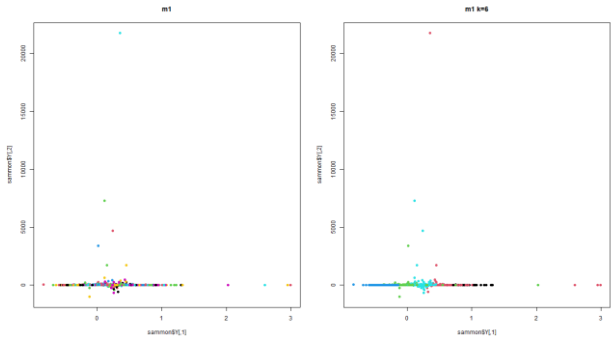


Figure 143 Sammon'smapping on Model 3

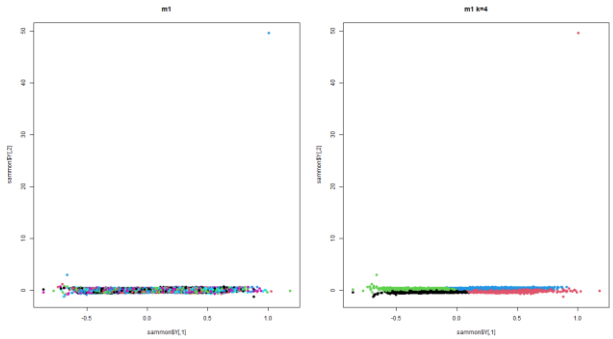


Figure 144 Sammon's mapping on Model 4

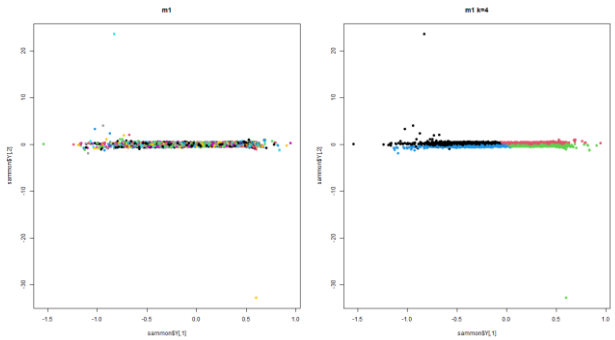


Figure 145 Sammon's mapping on Model 5

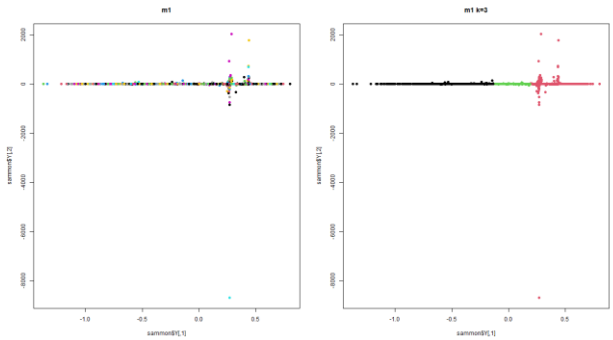


Figure 146 Sammon's mapping on Model 6

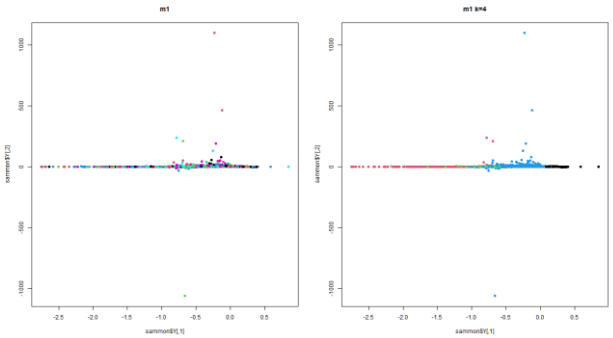


Figure 147 Sammon's mapping on Model 7

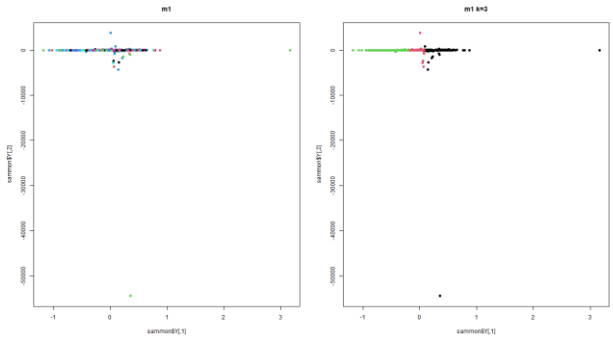


Figure 148 Sammon's mapping on Model 8

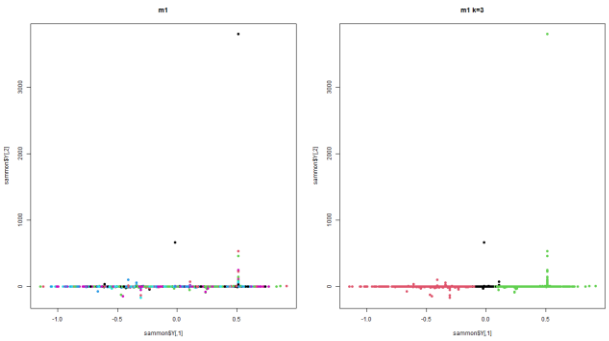


Figure 148 Sammon's mapping on Model 9

6.1.7 DBSCAN

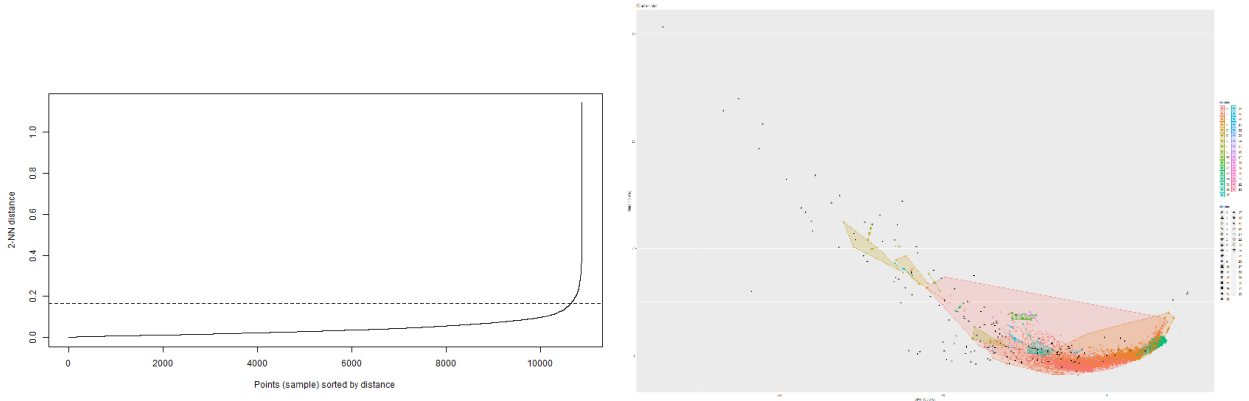


Figure 149 and 150 Knee plot and DBSCAN on Model 1

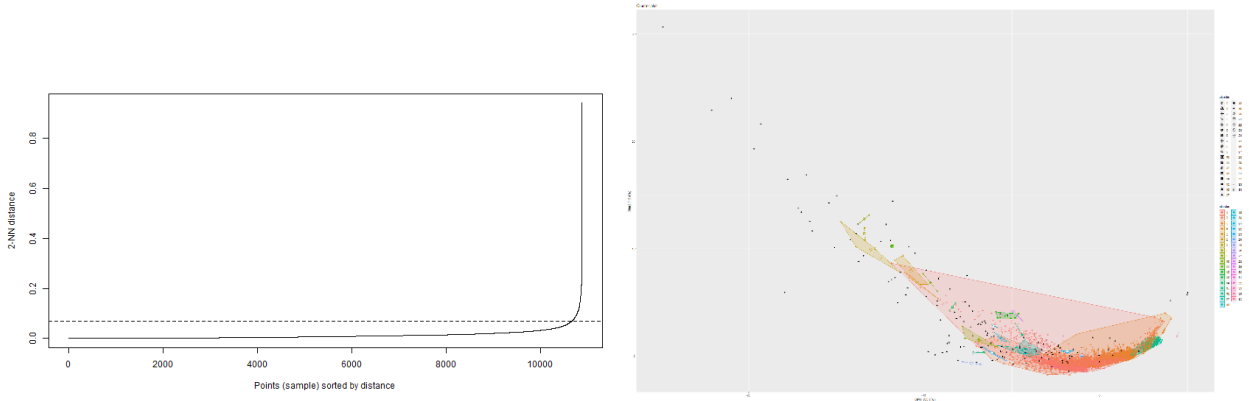


Figure 151 and 152 Knee plot and DBSCAN on Model 2

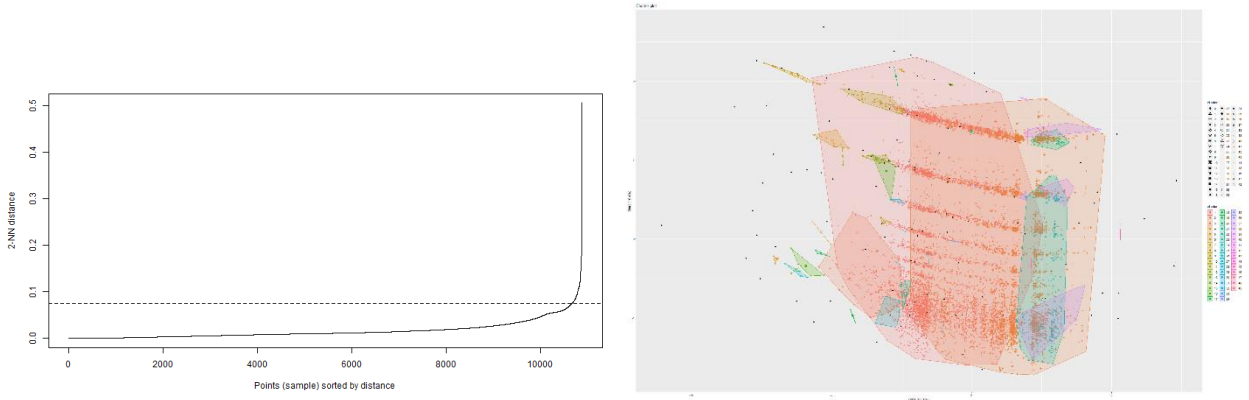


Figure 153 and 154 Knee plot and DBSCAN on Model 4

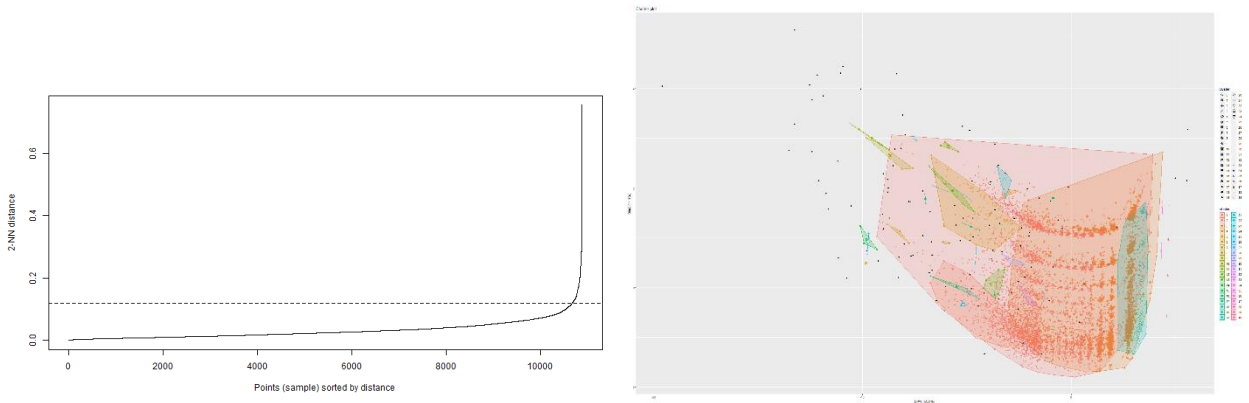


Figure 155 and 156 Knee plot and DBSCAN on Model 5

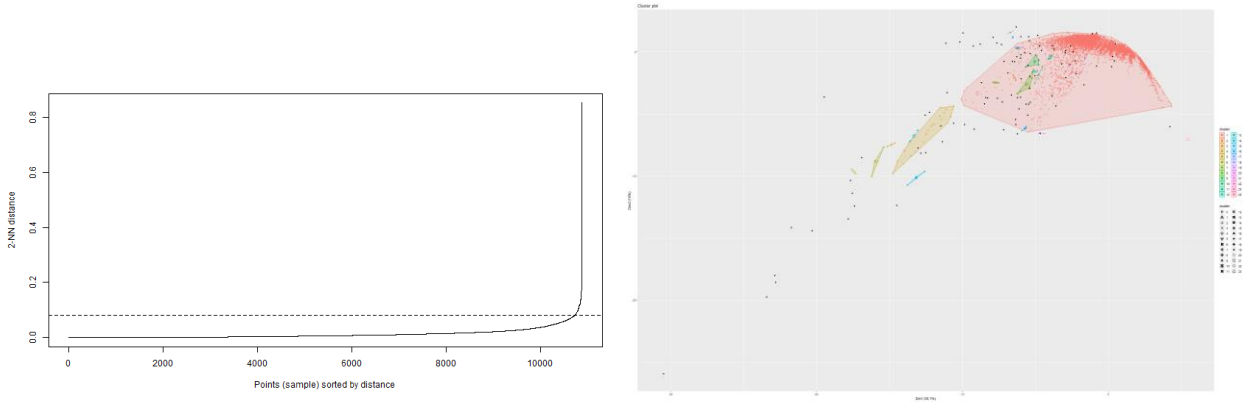


Figure 157 and 158 Knee plot and DBSCAN on Model 6

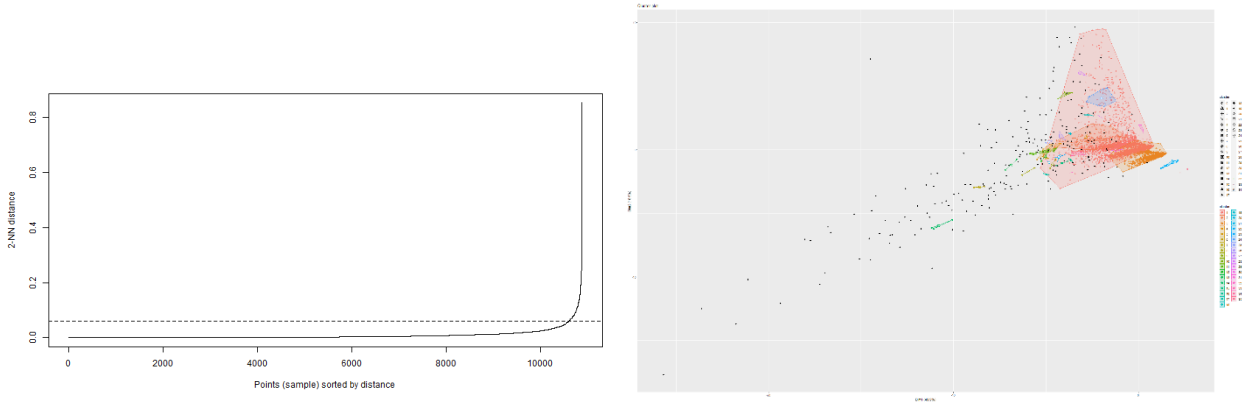


Figure 159 and 160 Knee plot and DBSCAN on Model 7

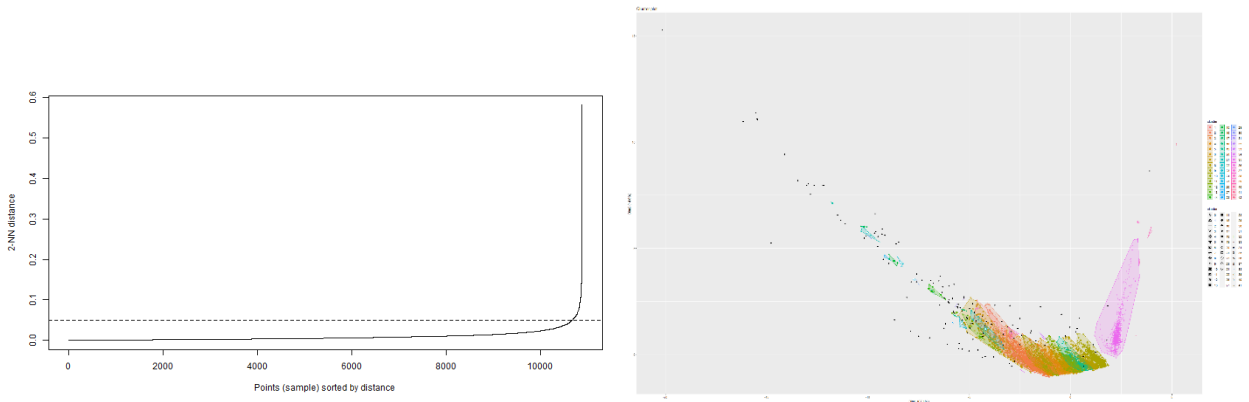


Figure 161 and 162 Knee plot and DBSCAN on Model 8

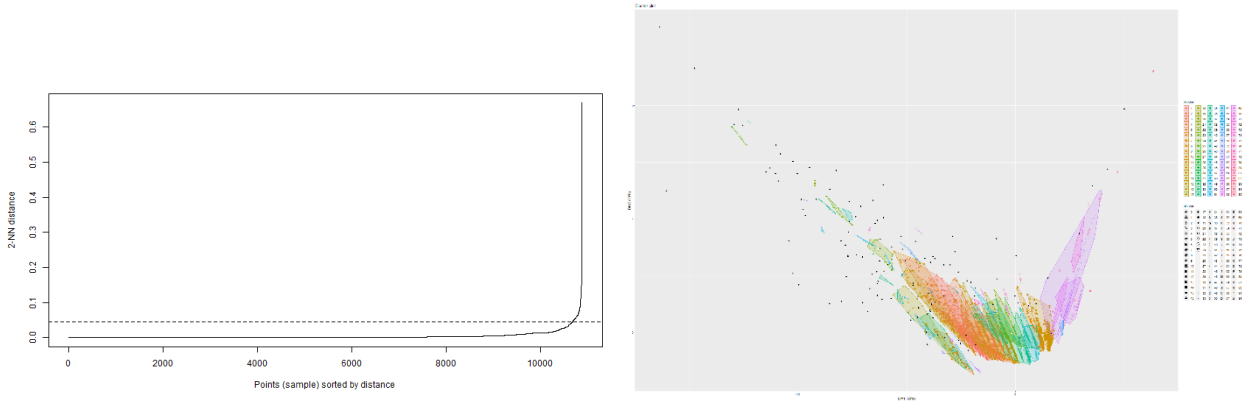


Figure 163 and 164 Knee plot and DBSCAN on Model 9

6.2 CA-HepPh dataset

6.2.1 Boxplot and Correlation plots

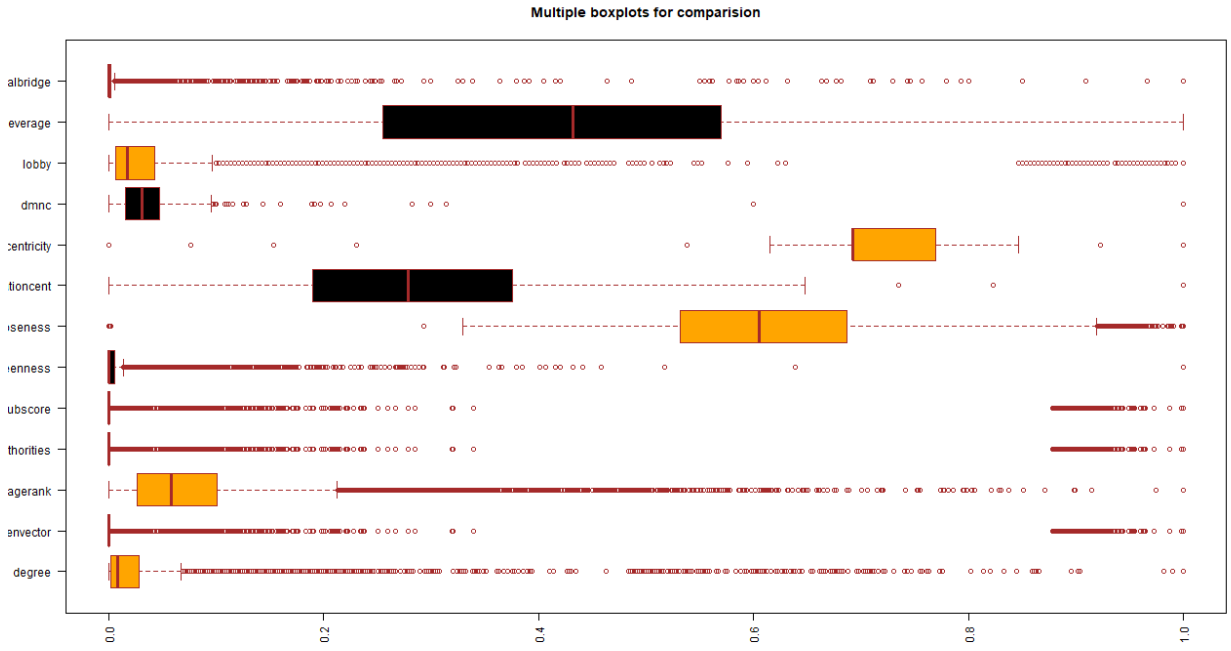


Figure 165 Boxplot

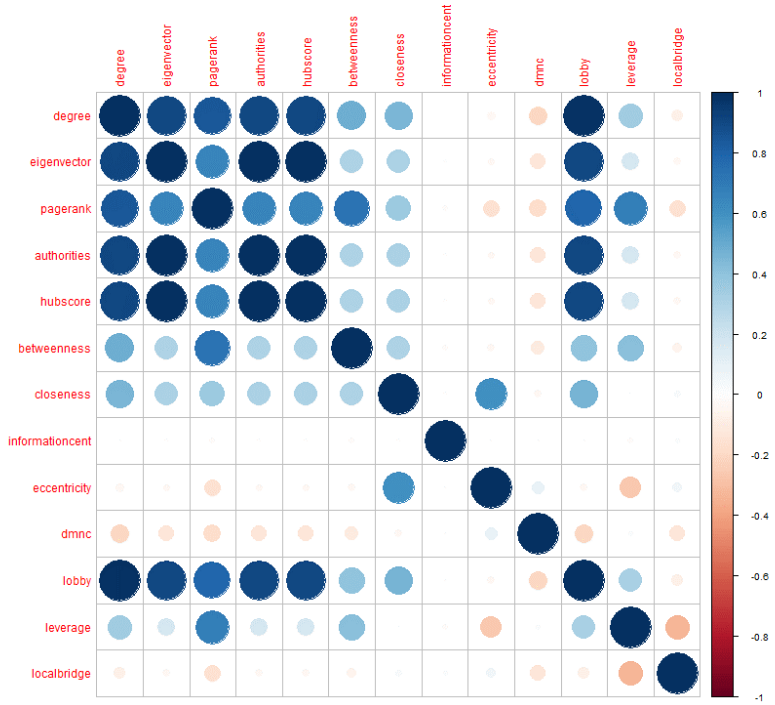


Figure 166 Correlation matrix

6.2.2 PCA

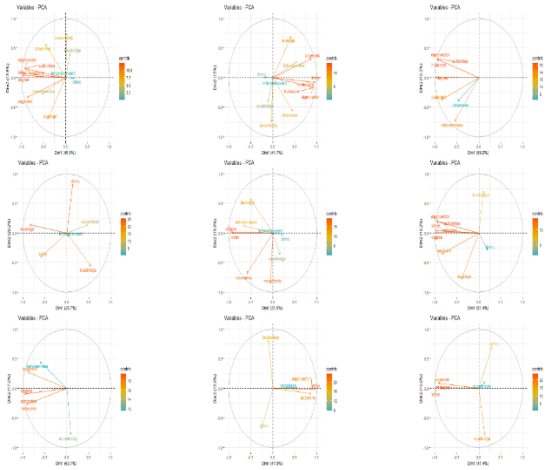


Figure 167 PCA

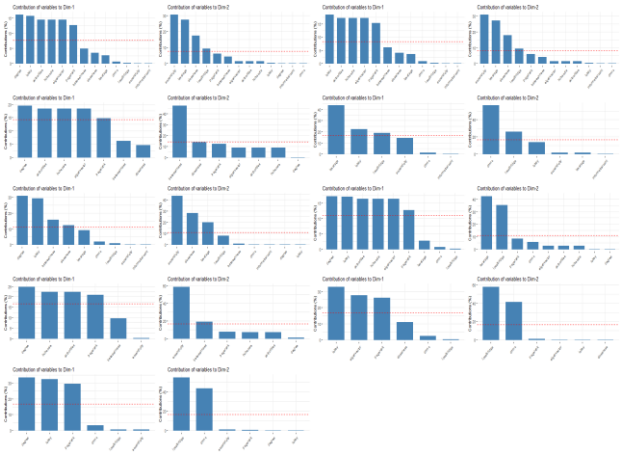


Figure 168 Dimension Contribution

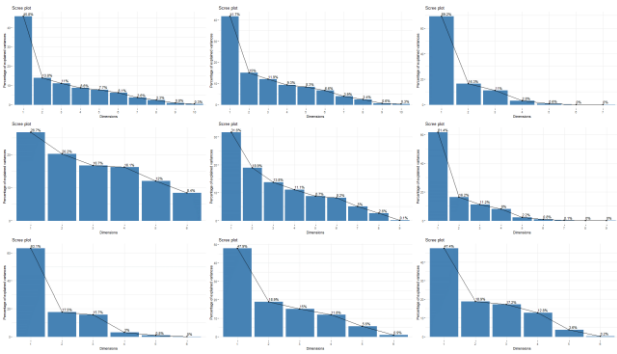


Figure 169 Scree plot

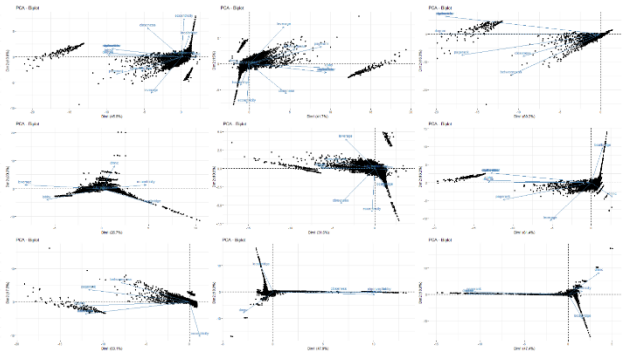


Figure 170 biplot

6.2.3 t-SNE and K-means on t-SNE

Model 1:

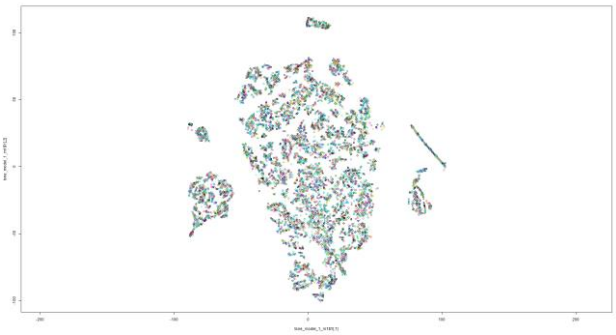


Figure 170 t-SNE on Model 1

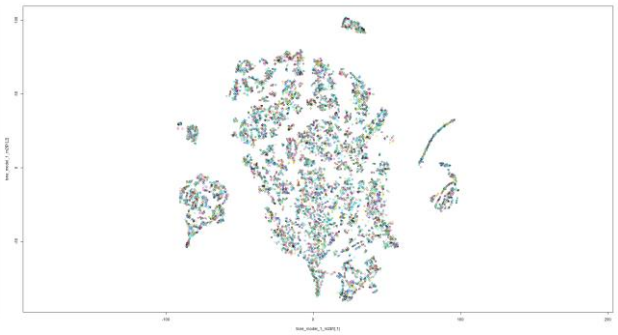


Figure 171 t-SNE on Model 2

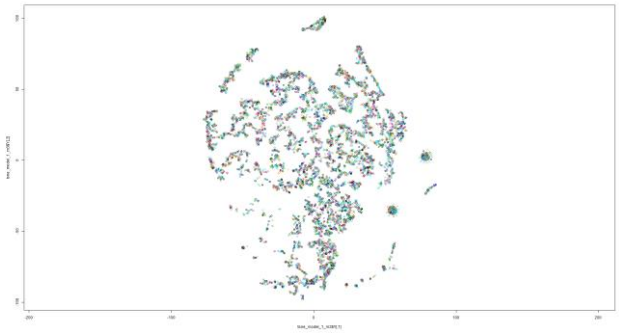


Figure 172 t-SNE on Model 3

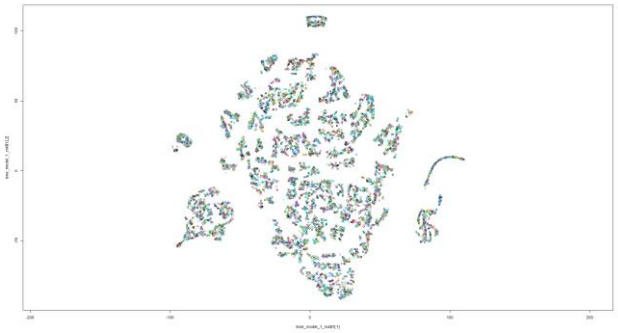


Figure 173 t-SNE on Model 4

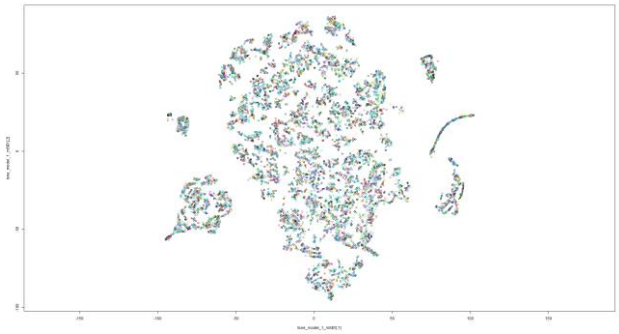


Figure 174 t-SNE on Model 5

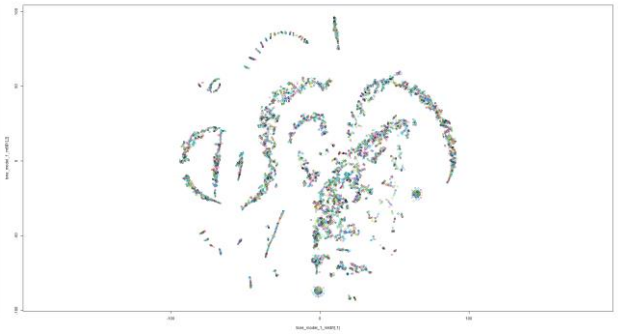


Figure 175 t-SNE on Model 6

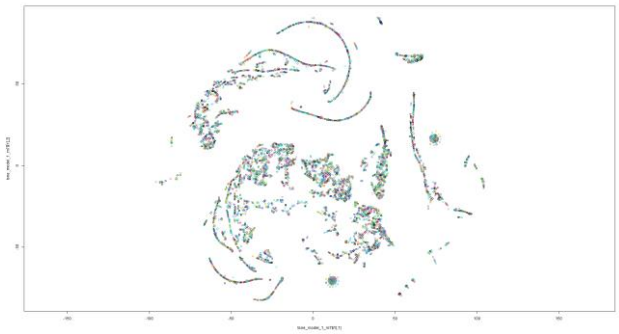


Figure 176 t-SNE on Model 7

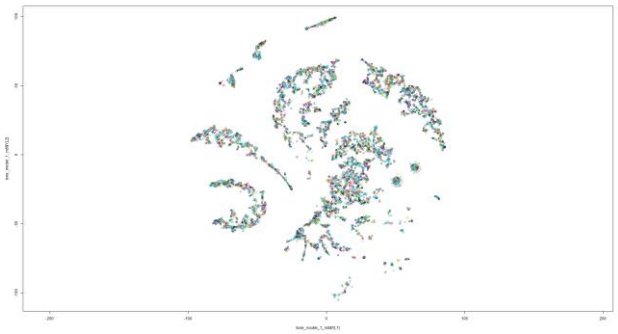


Figure 177 t-SNE on Model 8

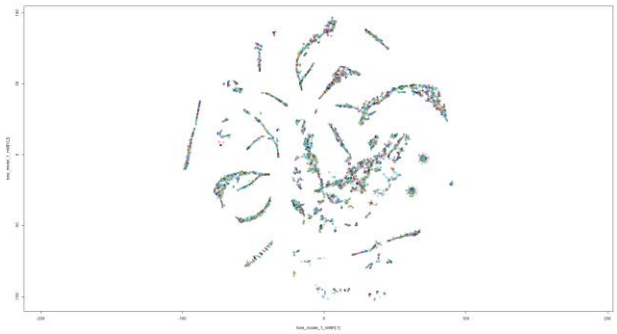


Figure 178 t-SNE on Model 9

Model 2:

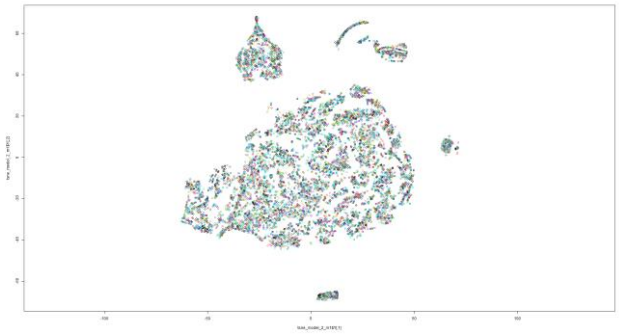


Figure 179 t-SNE on Model 1

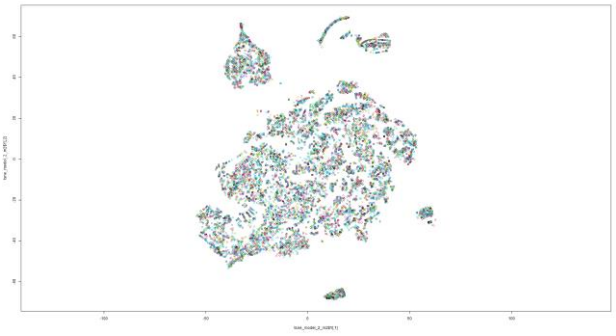


Figure 180 t-SNE on Model 2

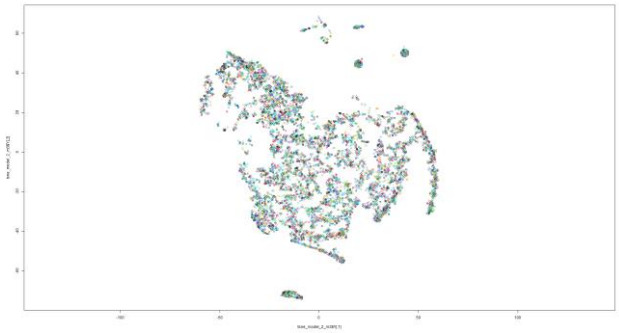


Figure 181 t-SNE on Model 3

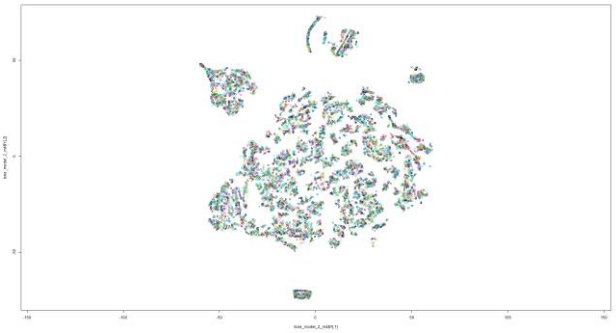


Figure 182 t-SNE on Model 4

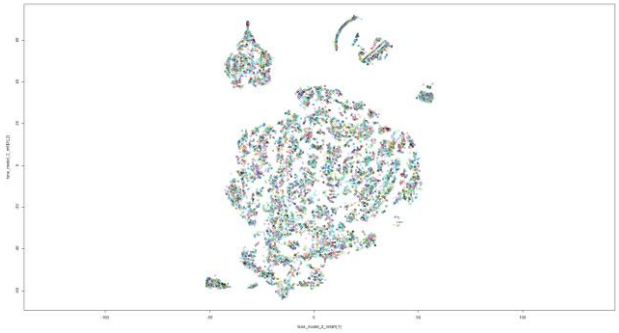


Figure 183 t-SNE on Model 5

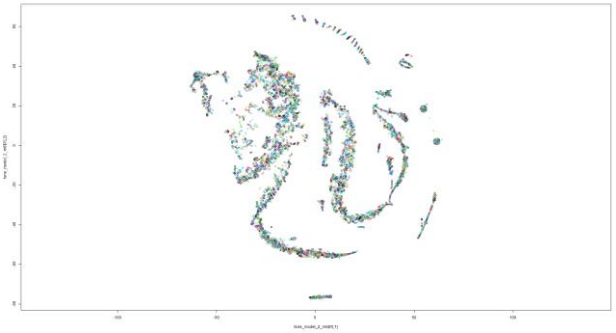


Figure 184 t-SNE on Model 6

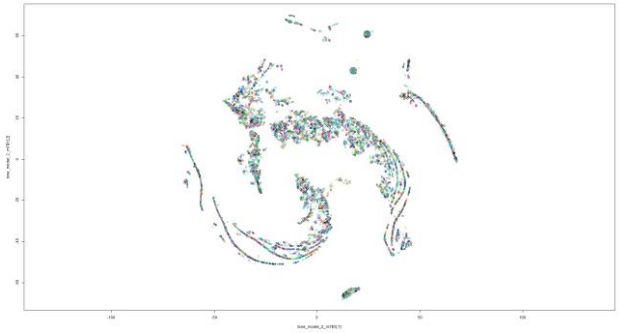


Figure 185 t-SNE on Model 7

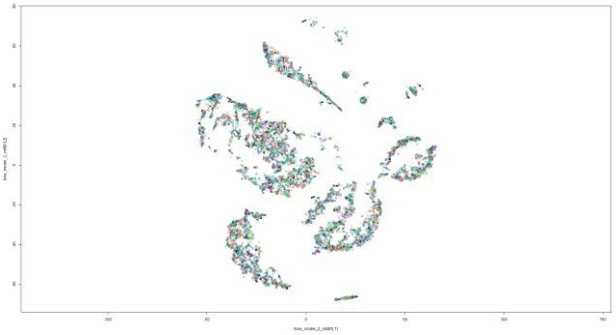


Figure 186 t-SNE on Model 8

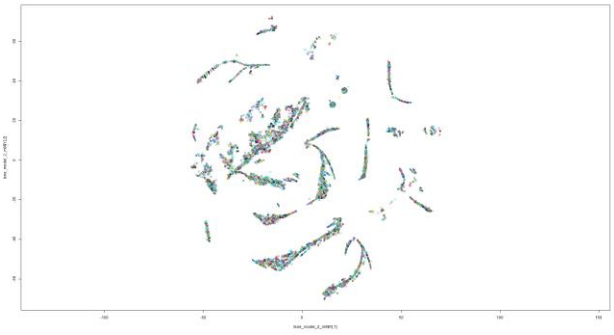


Figure 187 t-SNE on Model 9

Model 3:

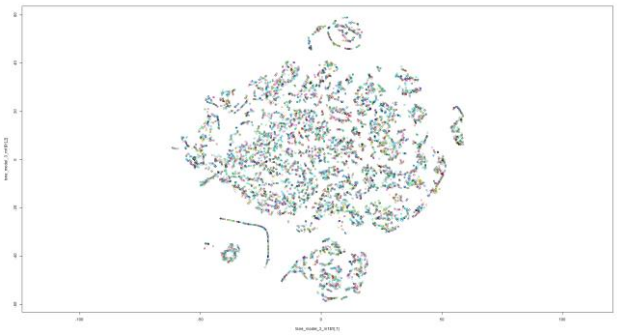


Figure 188 t-SNE on Model 1

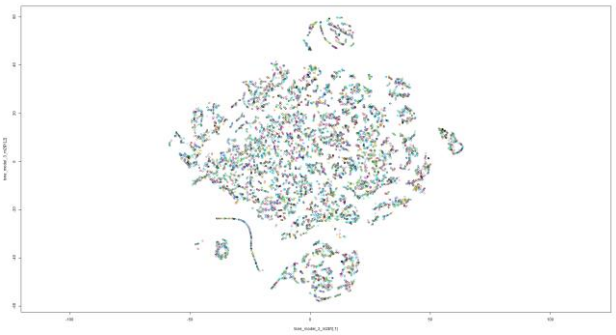


Figure 189 t-SNE on Model 2

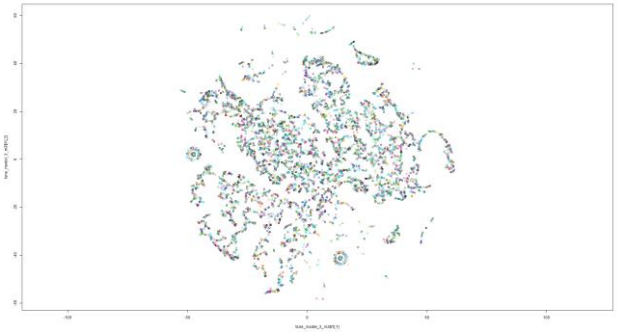


Figure 190 t-SNE on Model 3

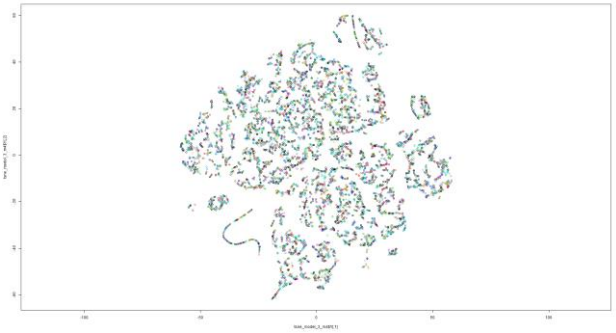


Figure 191 t-SNE on Model 4

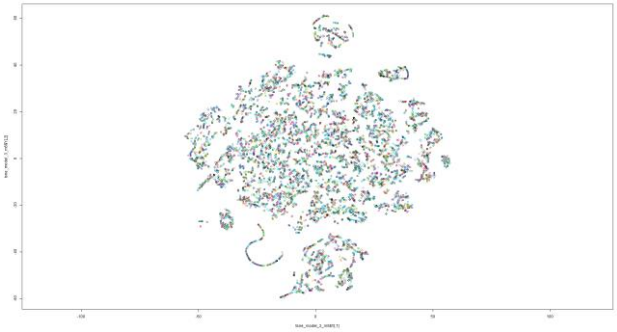


Figure 192 t-SNE on Model 5

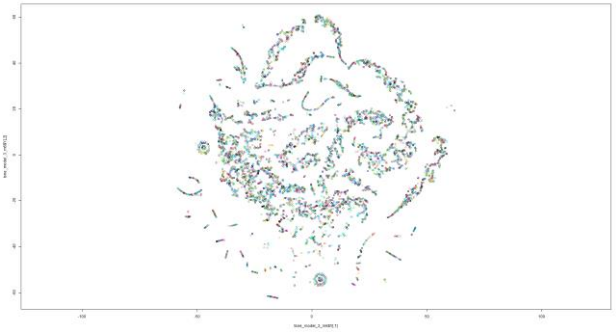


Figure 193 t-SNE on Model 6

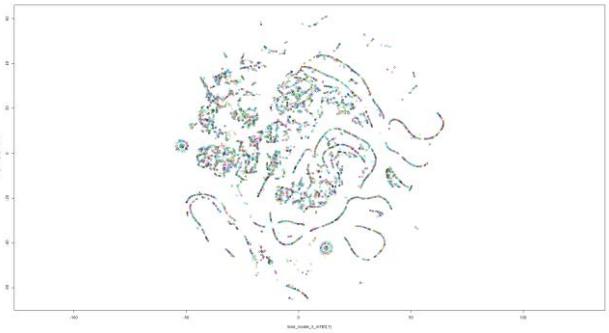


Figure 194 t-SNE on Model 7

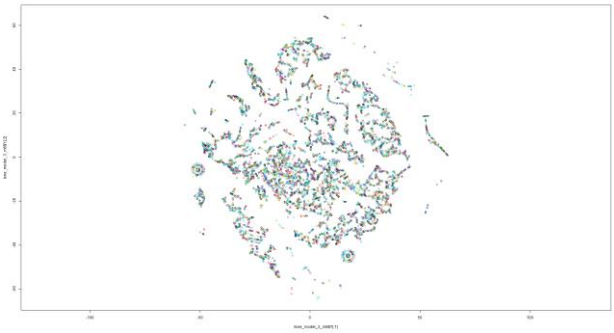


Figure 195 t-SNE on Model 8

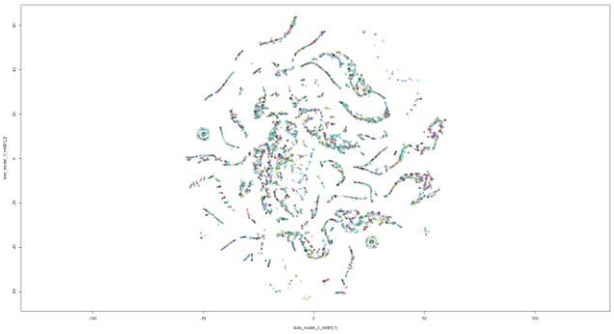


Figure 196 t-SNE on Model 9

Model 4:

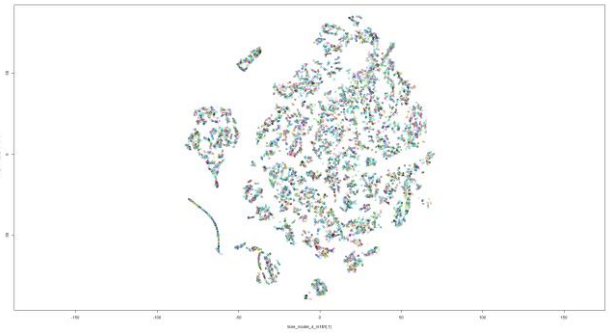


Figure 197 t-SNE on Model 1

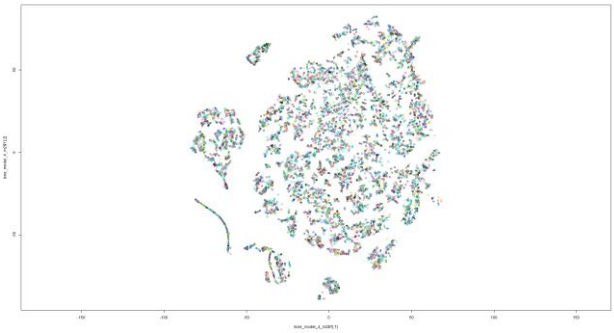


Figure 198 t-SNE on Model 2

Figure 199 t-SNE on Model 3

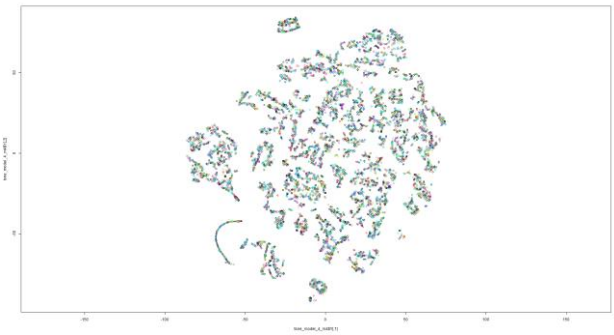


Figure 200 t-SNE on Model 4

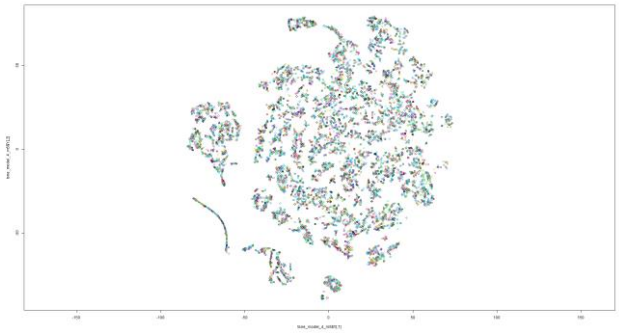


Figure 201 t-SNE on Model 5

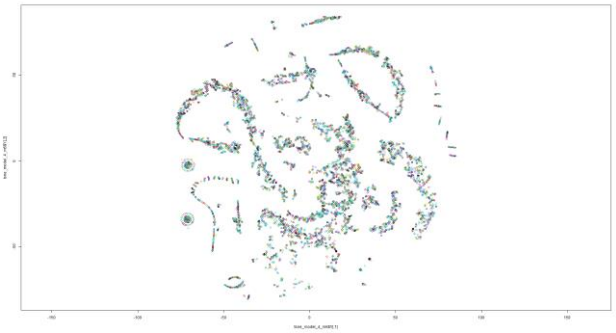


Figure 202 t-SNE on Model 6

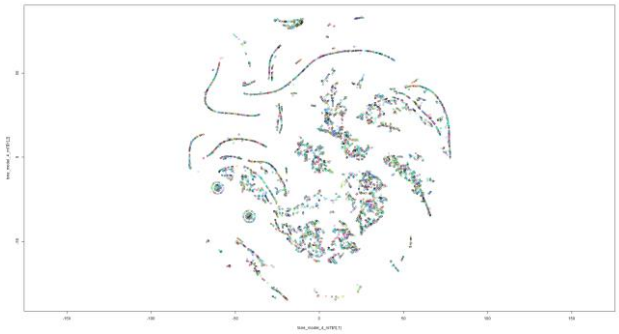


Figure 203 t-SNE on Model 7

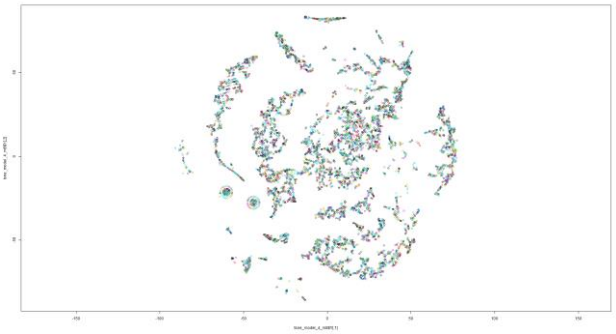


Figure 204 t-SNE on Model 8

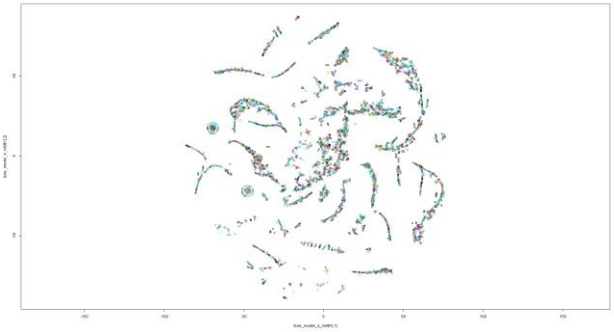


Figure 205 t-SNE on Model 9

K-means on t-SNE:

Model 1:

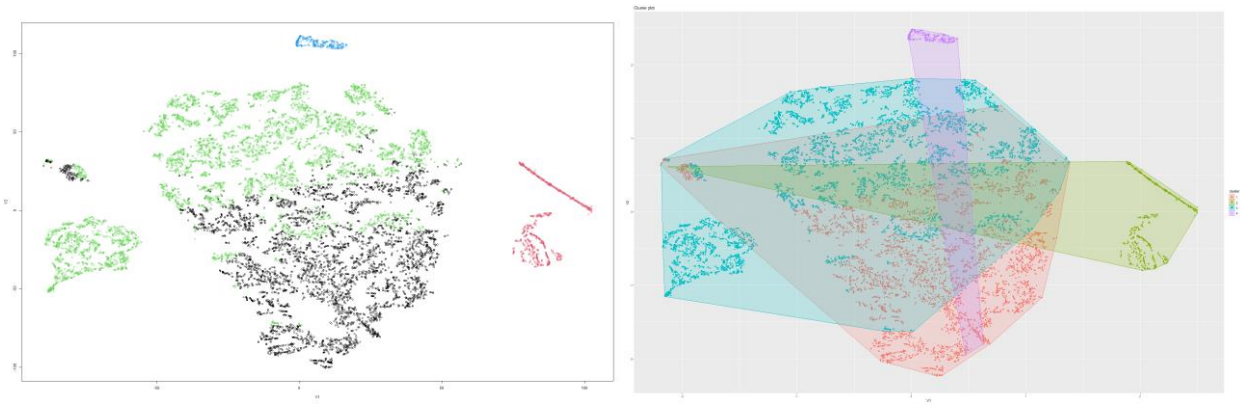


Figure 206&207 k-Means on t-SNE Model 1

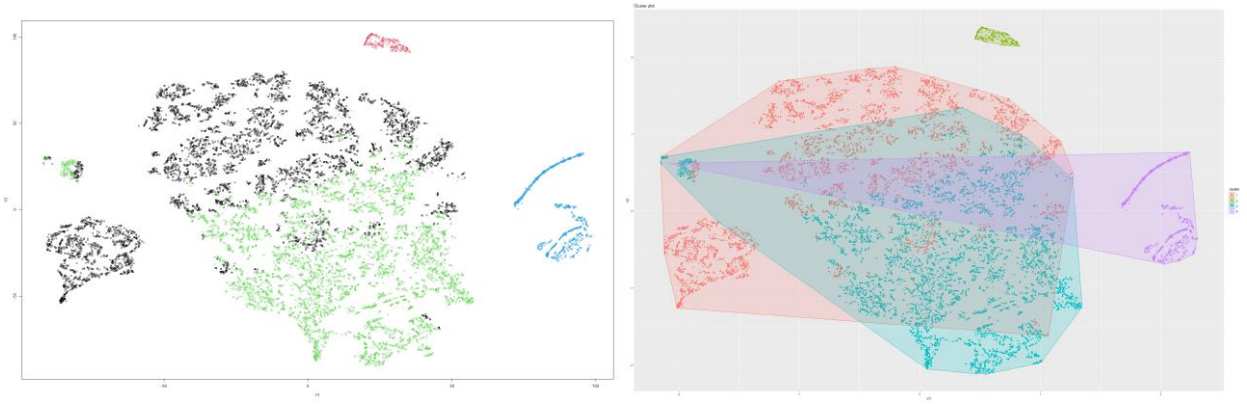


Figure 208&209 k-Means on t-SNE Model 2

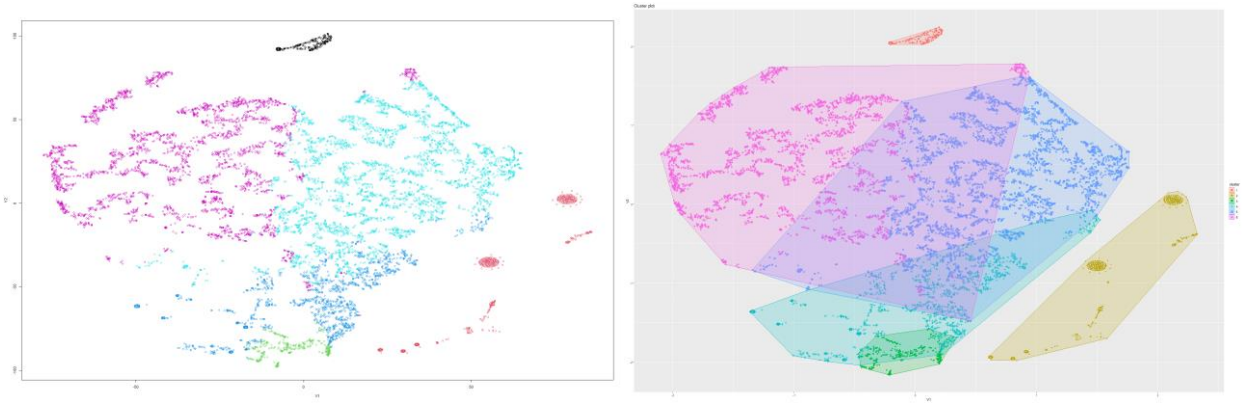


Figure 210&211 k-Means on t-SNE Model 3

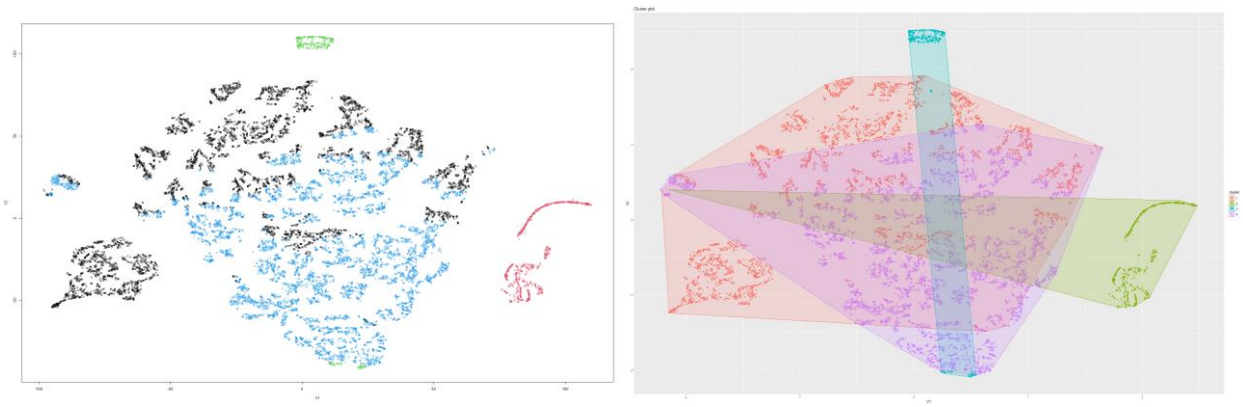


Figure 212&213 k-Means on t-SNE Model 4

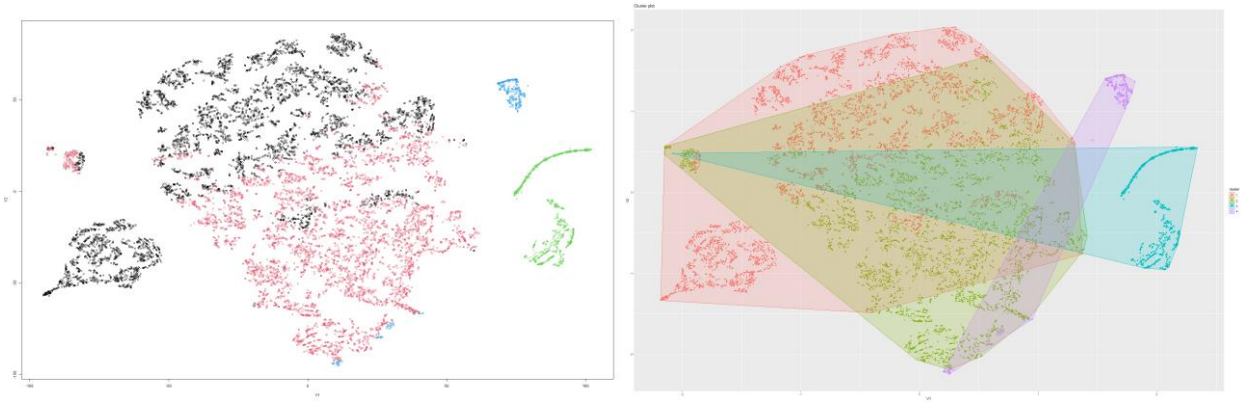


Figure 214&215 k-Means on t-SNE Model 5

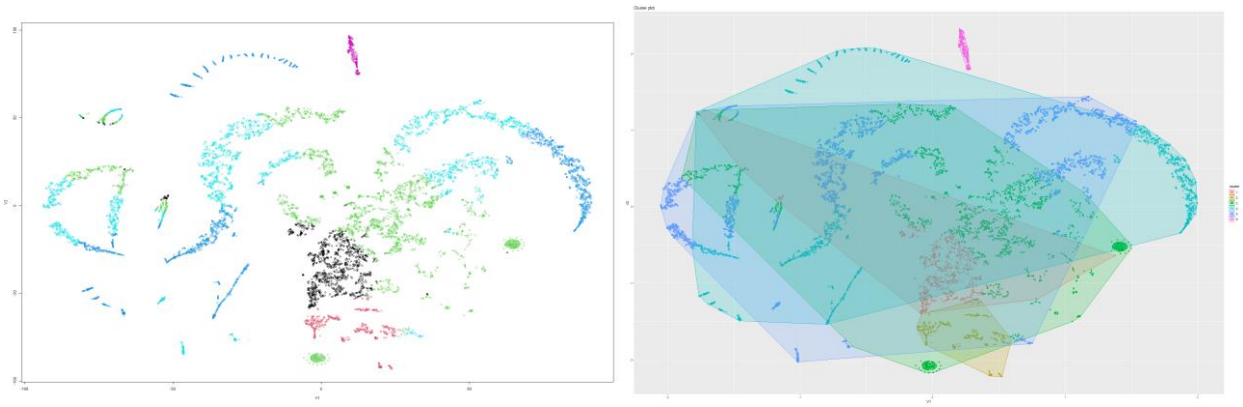


Figure 216&217 k-Means on t-SNE Model 6

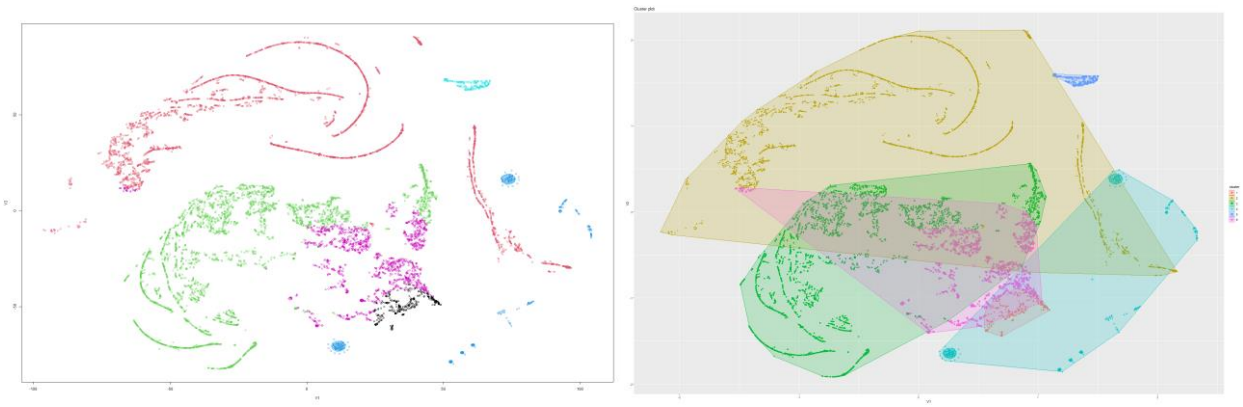


Figure 218&219 k-Means on t-SNE Model 7

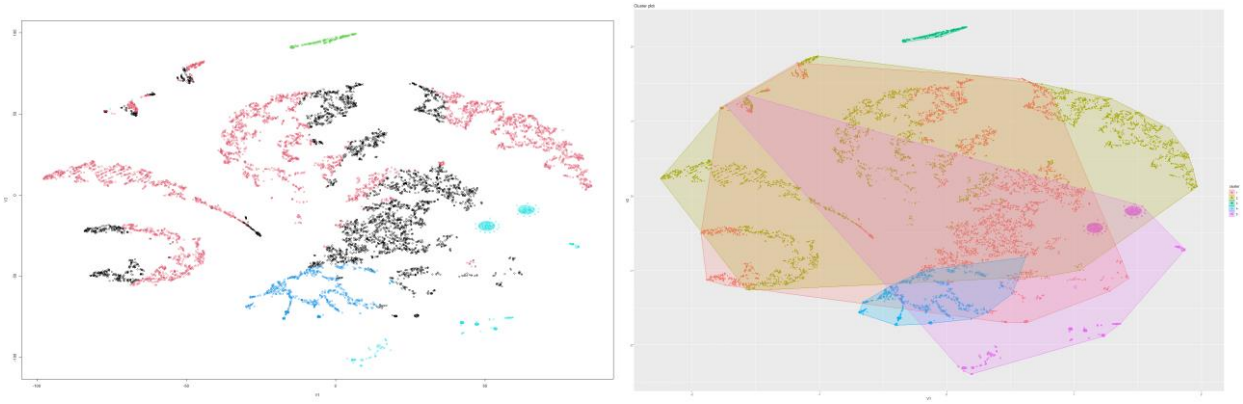


Figure 220&221 k-Means on t-SNE Model 8

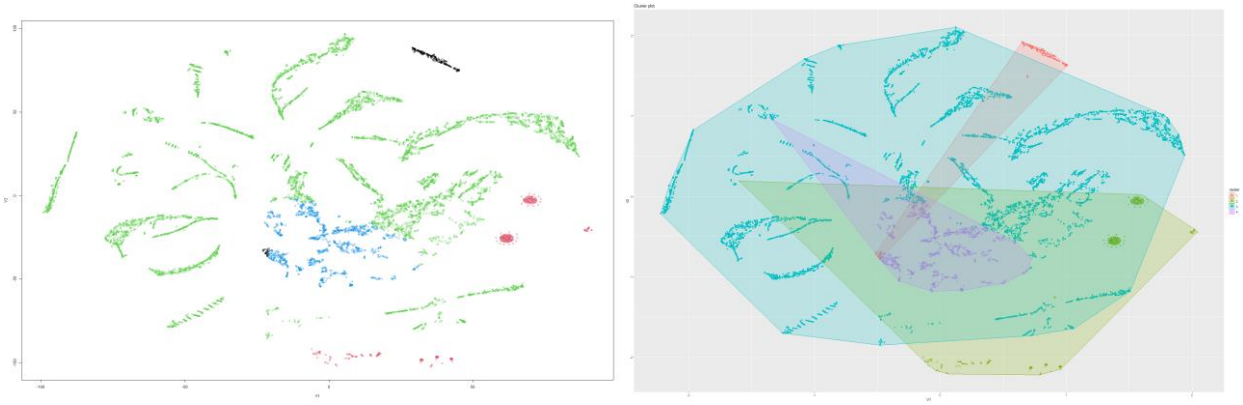


Figure 222&223 k-Means on t-SNE Model 9

Model 2:

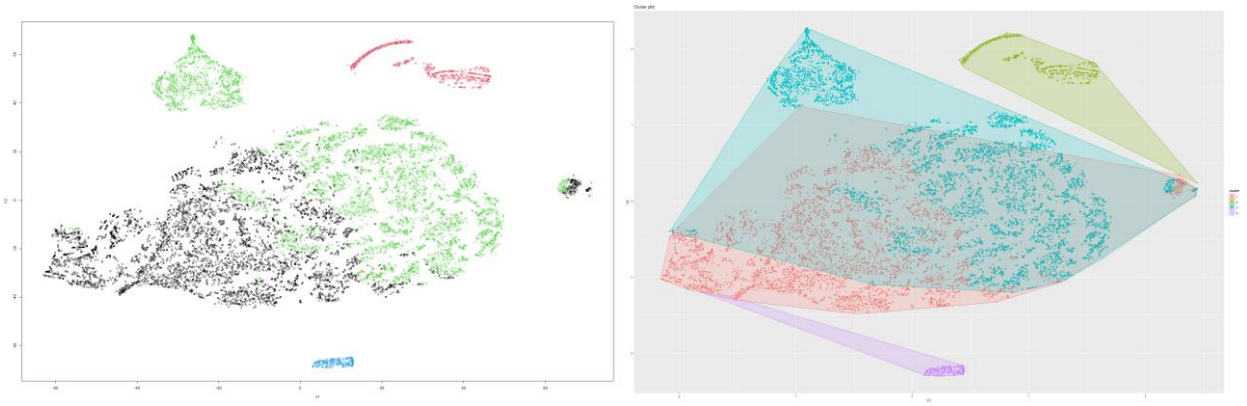


Figure 224&225 k-Means on t-SNE Model 1

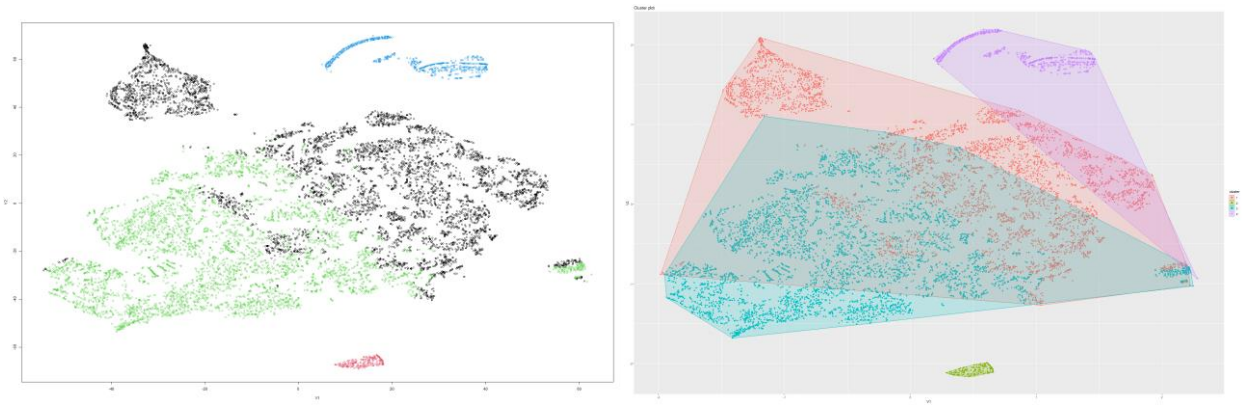


Figure 226&227 k-Means on t-SNE Model 2

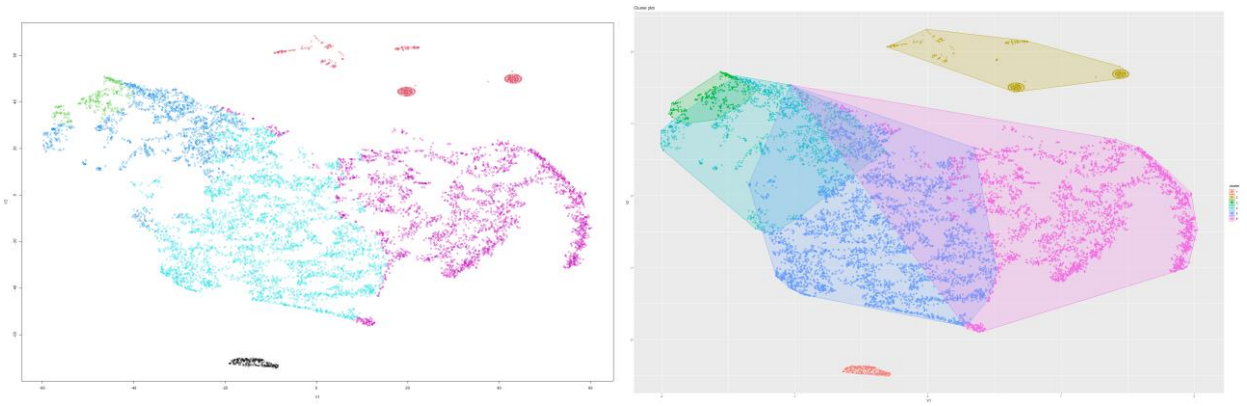


Figure 228&229 k-Means on t-SNE Model 3

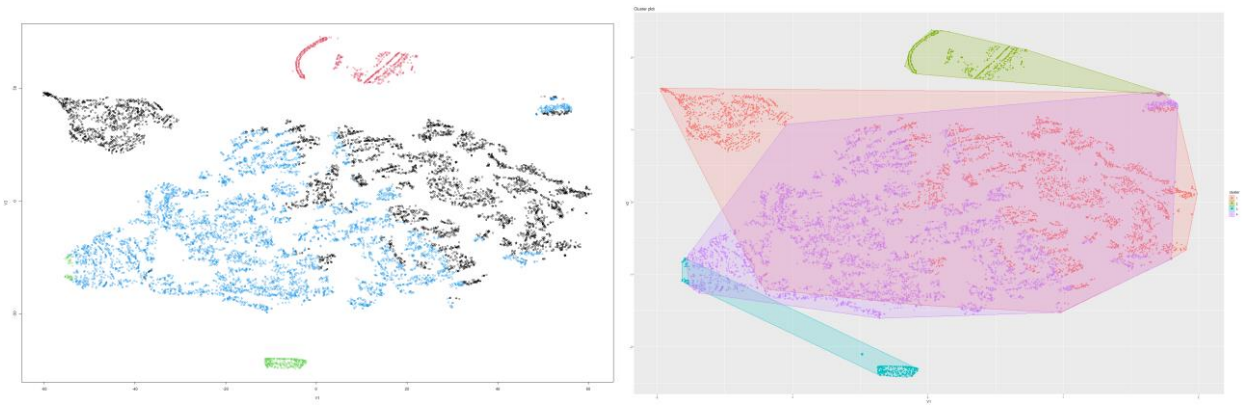


Figure 230&231 k-Means on t-SNE Model 4

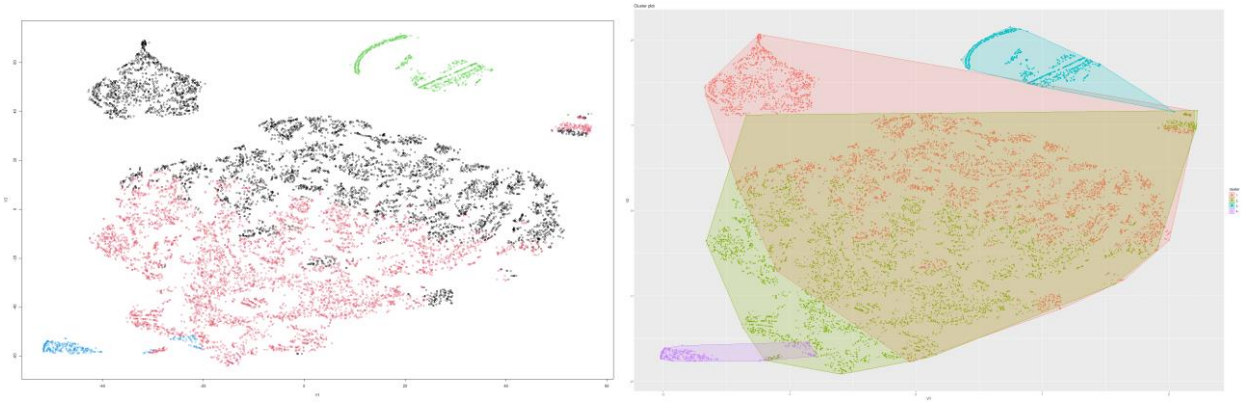


Figure 232&233 k-Means on t-SNE Model 5

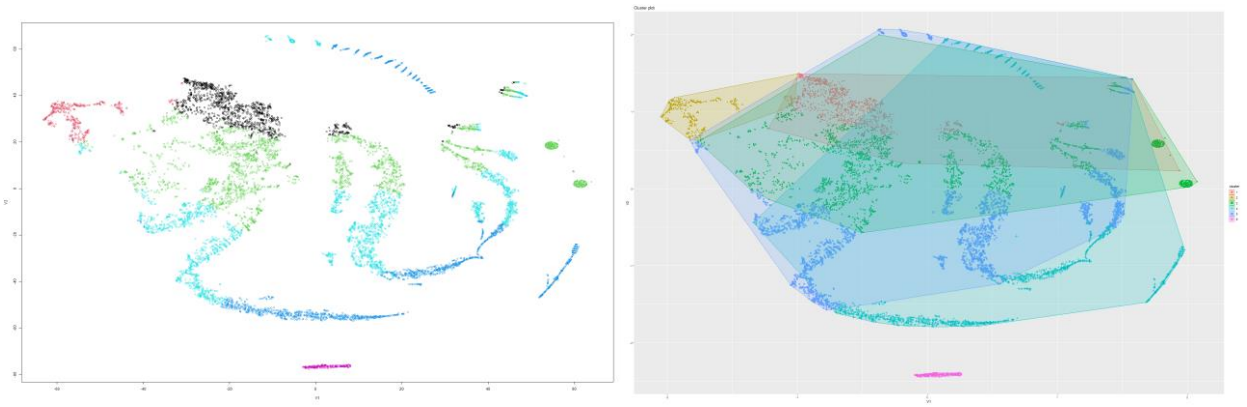


Figure 234&235 k-Means on t-SNE Model 6

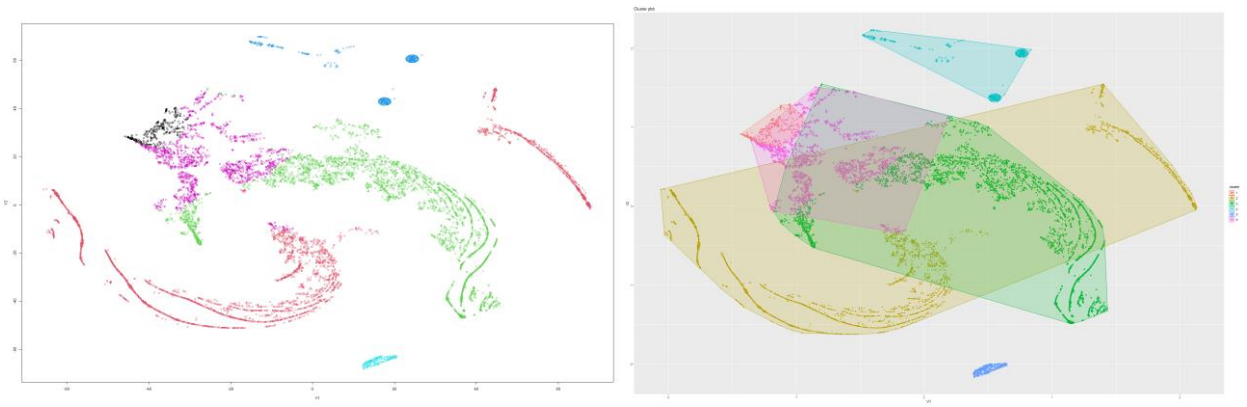


Figure 236&237 k-Means on t-SNE Model 7

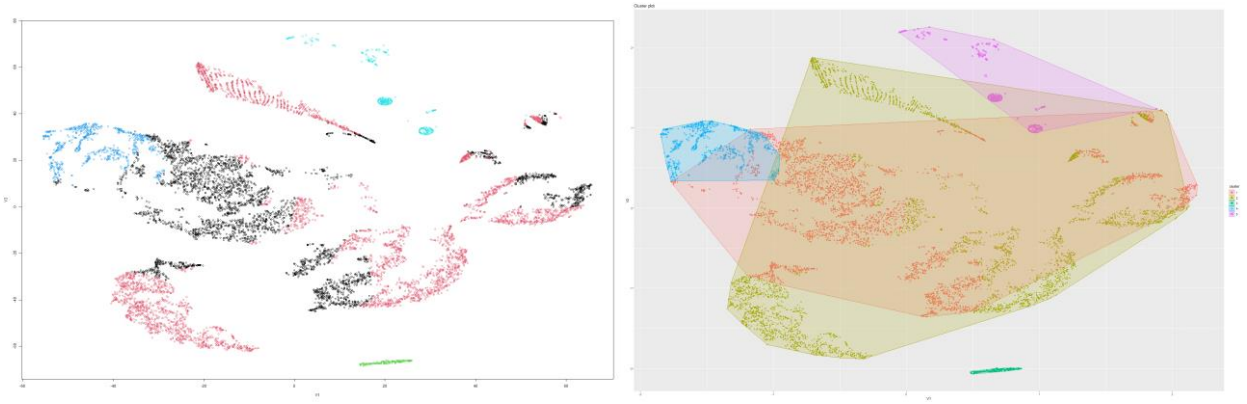


Figure 238&239 k-Means on t-SNE Model 8

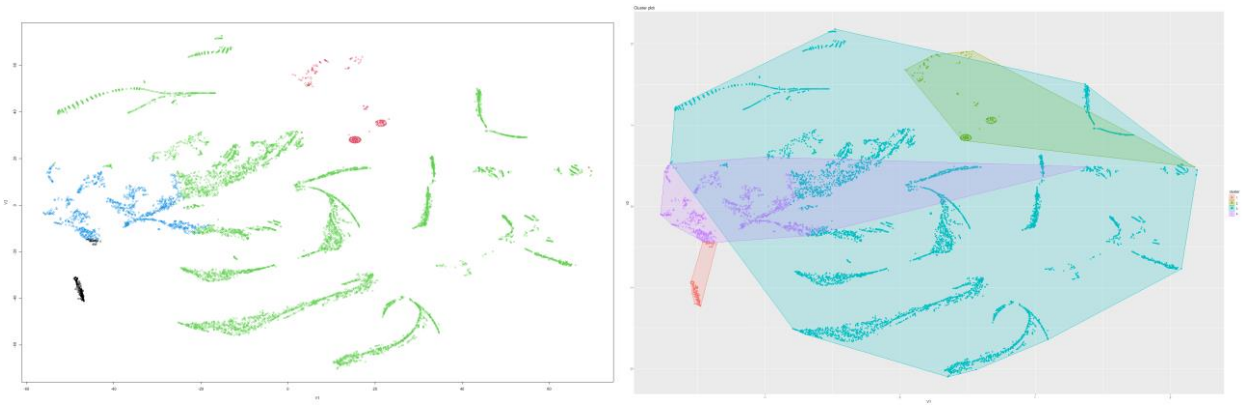


Figure 240&241 k-Means on t-SNE Model 9

Model 3:

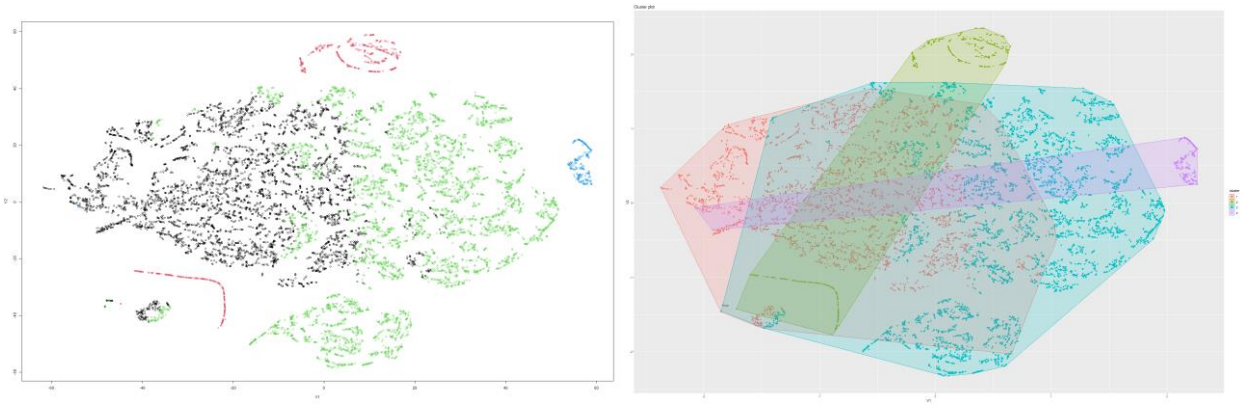


Figure 242&243 k-Means on t-SNE Model 1

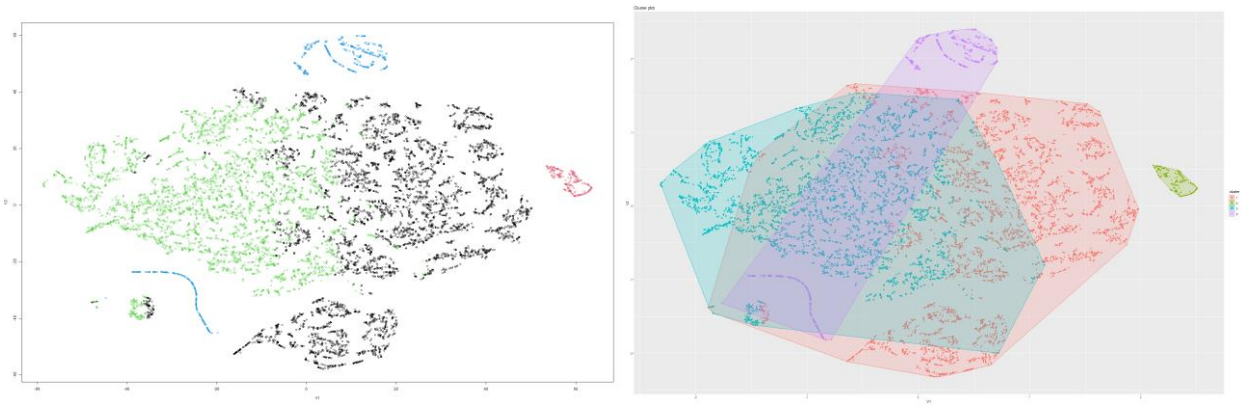


Figure 244&245 k-Means on t-SNE Model 2

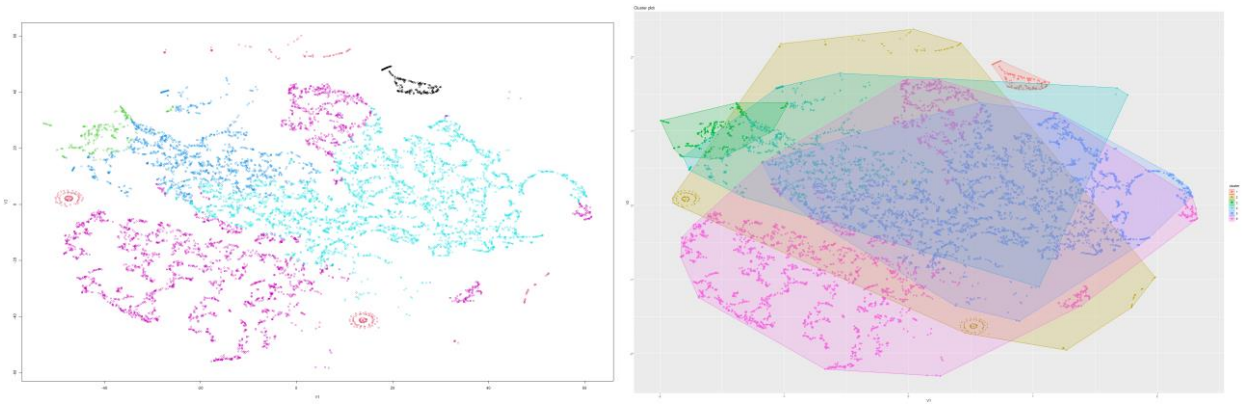


Figure 246&247 k-Means on t-SNE Model 3

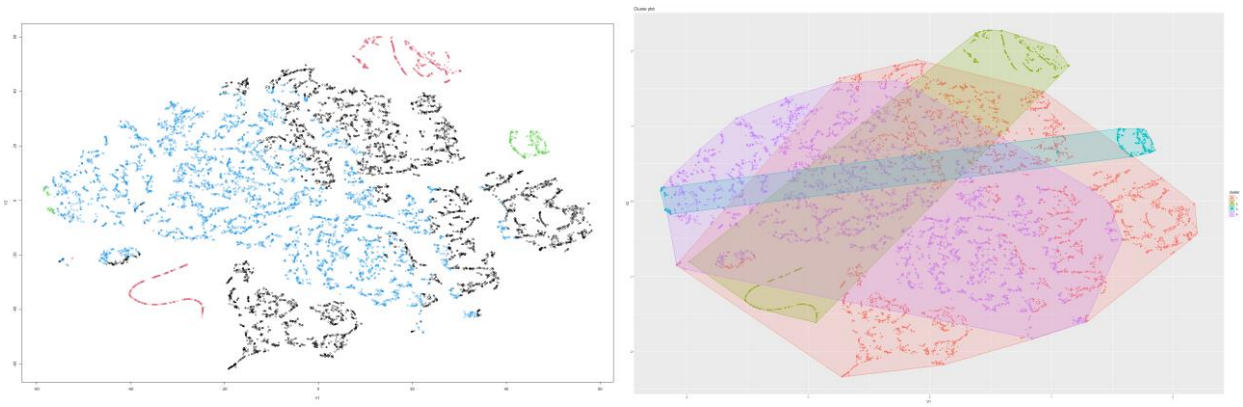


Figure 248&249 k-Means on t-SNE Model 4

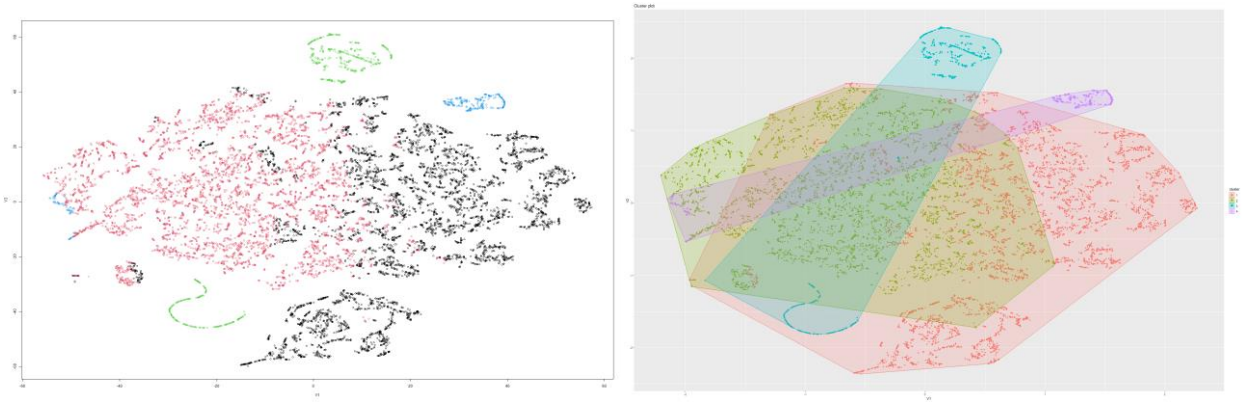


Figure 250&251 k-Means on t-SNE Model 5

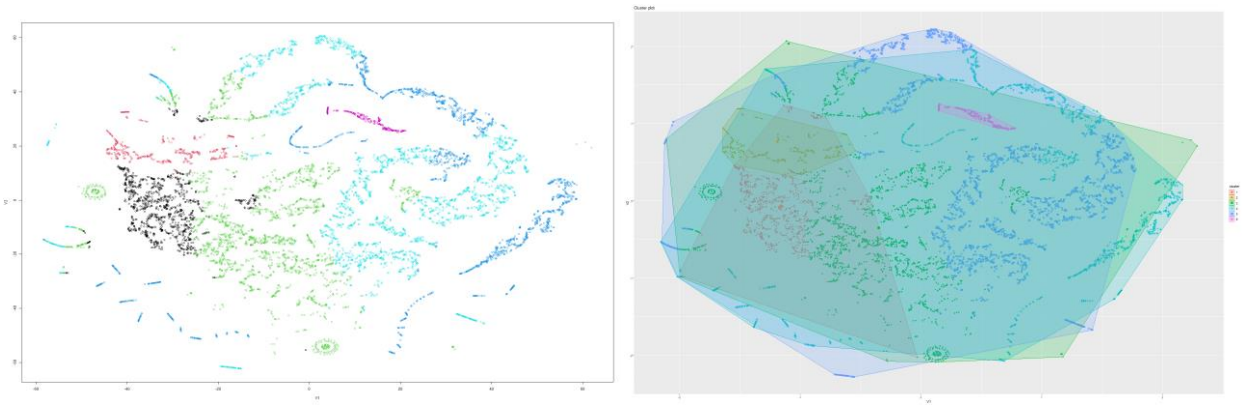


Figure 252&253 k-Means on t-SNE Model 6

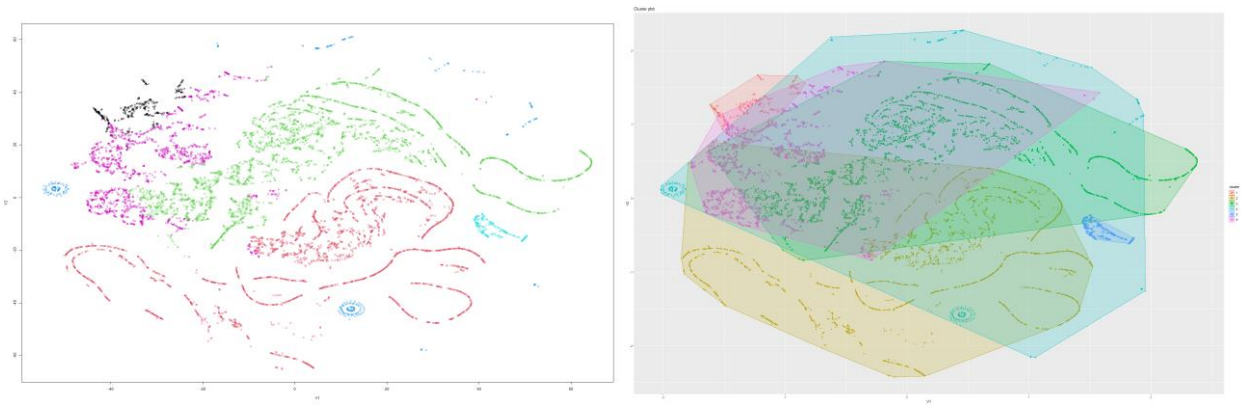


Figure 254&255 k-Means on t-SNE Model 7

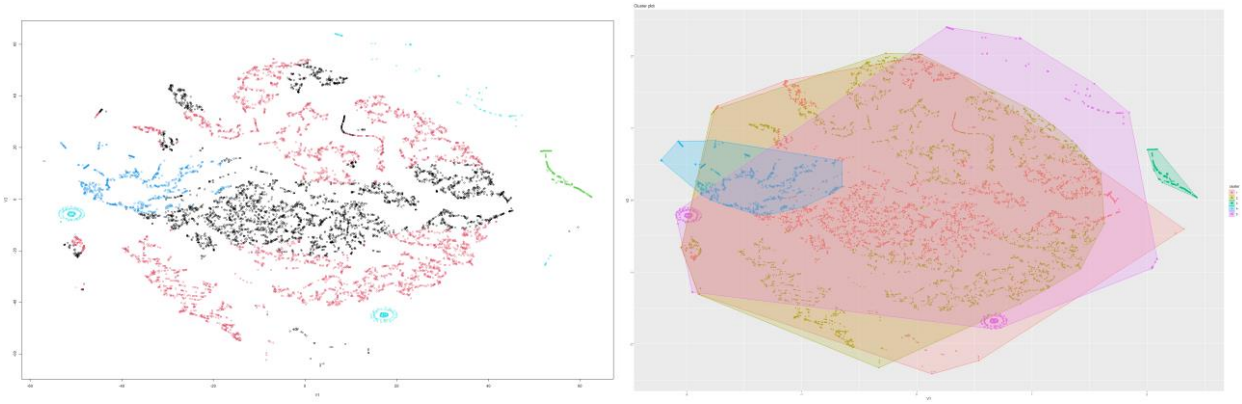


Figure 256&257 k-Means on t-SNE Model 8

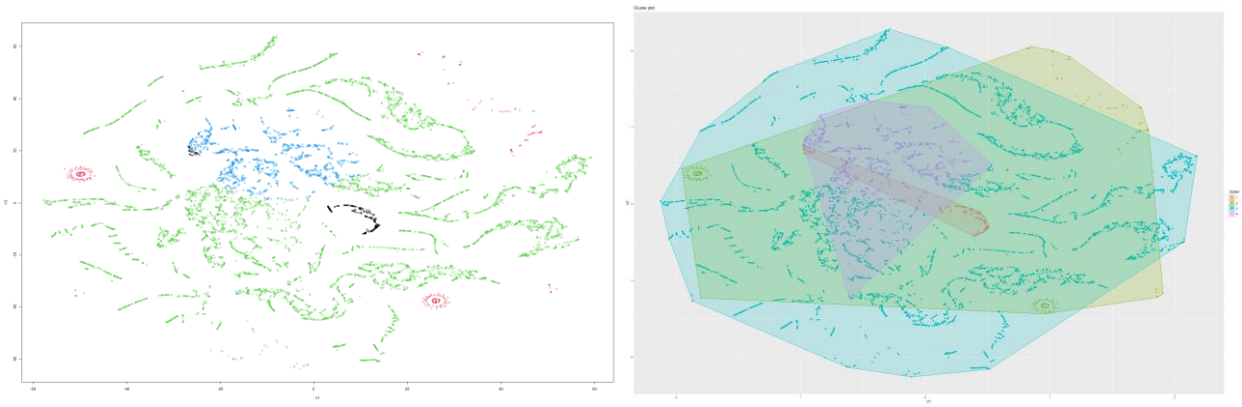


Figure 258&259 k-Means on t-SNE Model 9

Model 4:

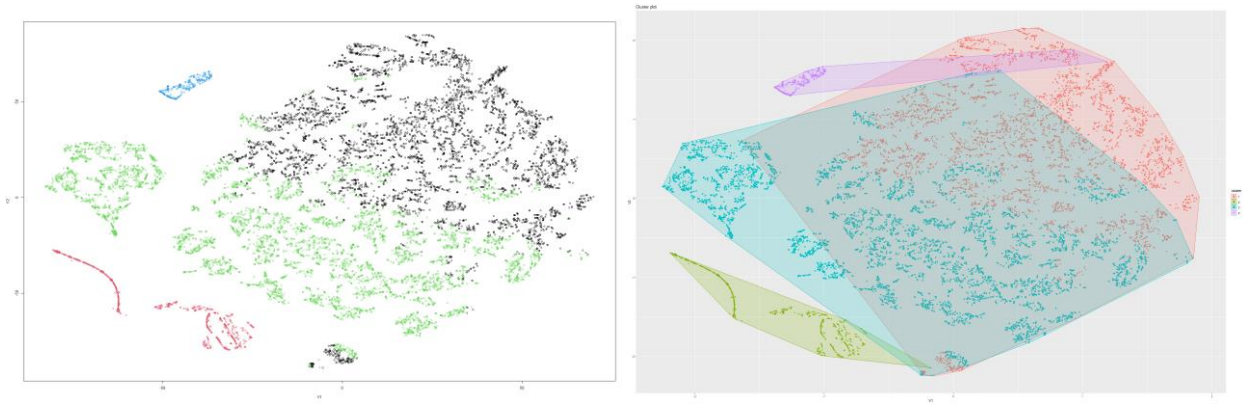


Figure 260&261 k-Means on t-SNE Model 1

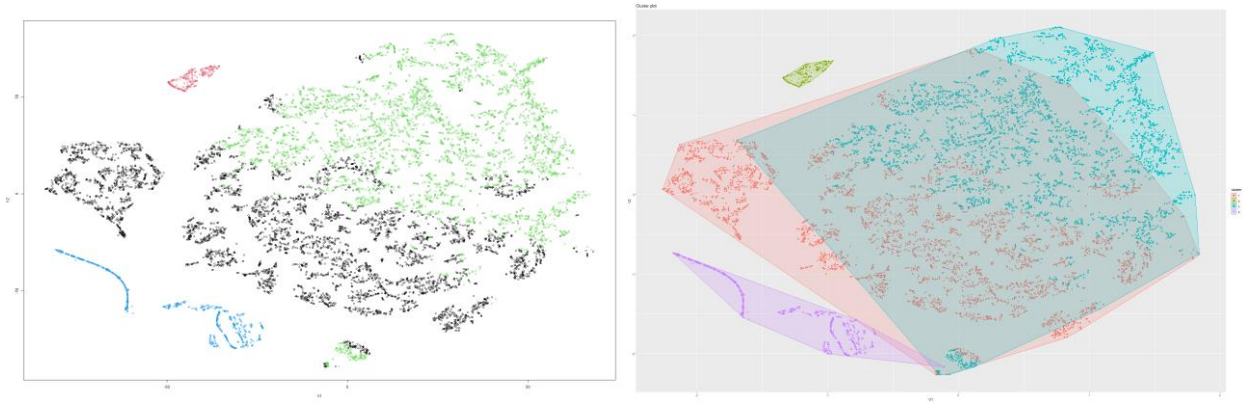


Figure 262&263 k-Means on t-SNE Model 2

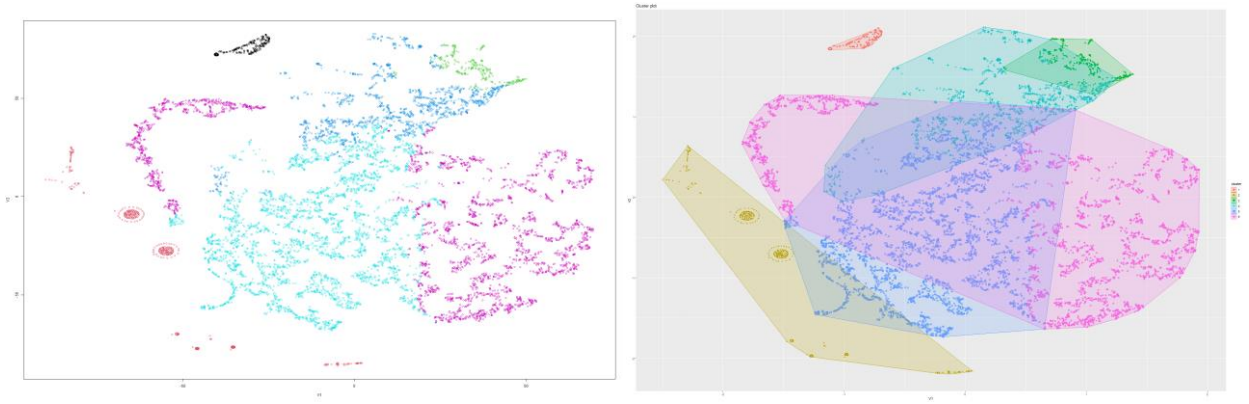


Figure 264&265 k-Means on t-SNE Model 3

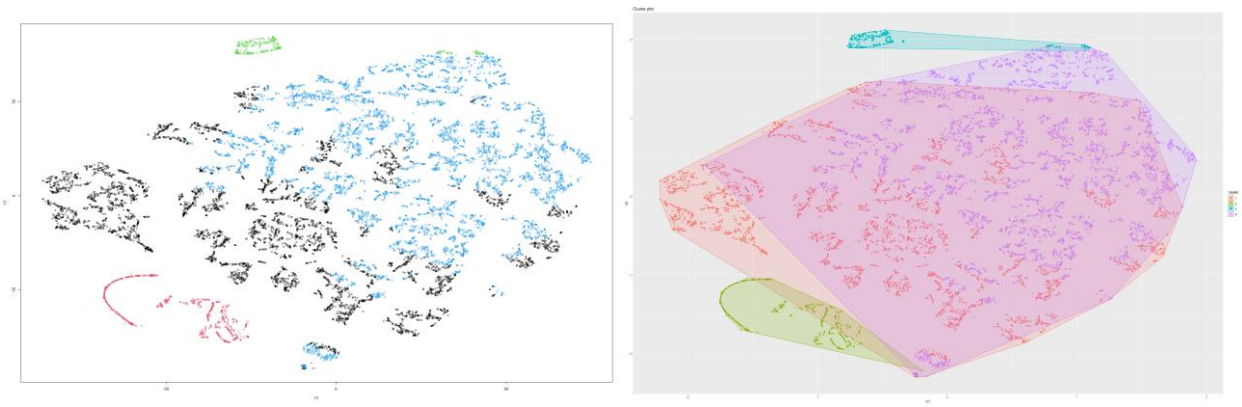


Figure 266&267 k-Means on t-SNE Model 4

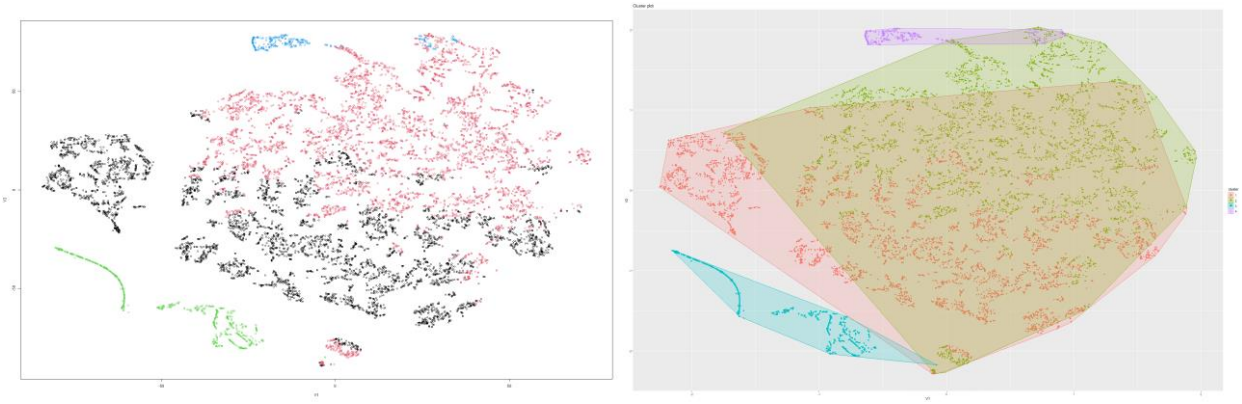


Figure 268&269 k-Means on t-SNE Model 5

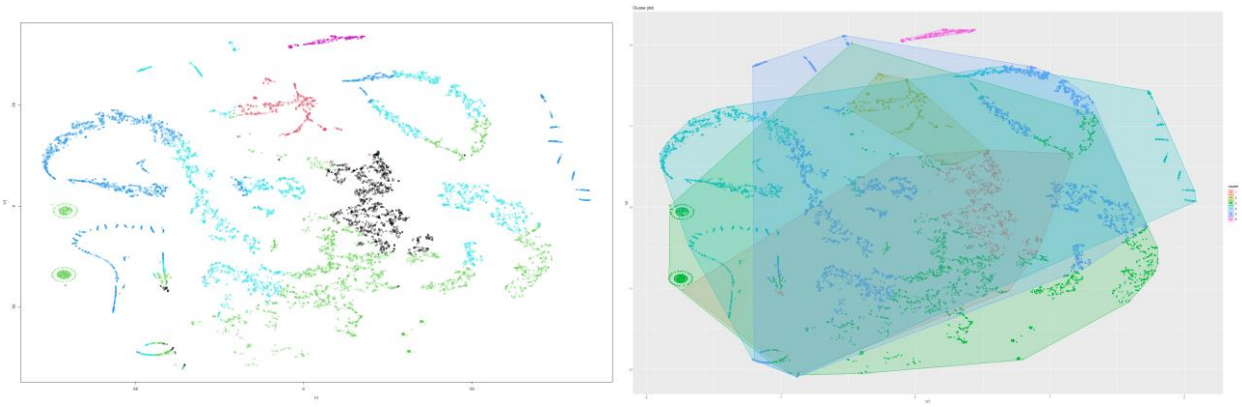


Figure 270&271 k-Means on t-SNE Model 6

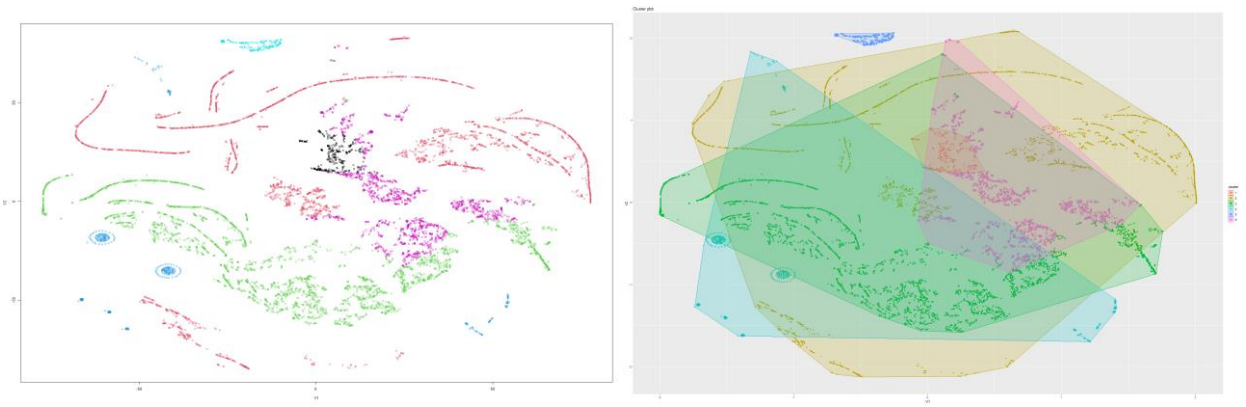


Figure 272&273 k-Means on t-SNE Model 7

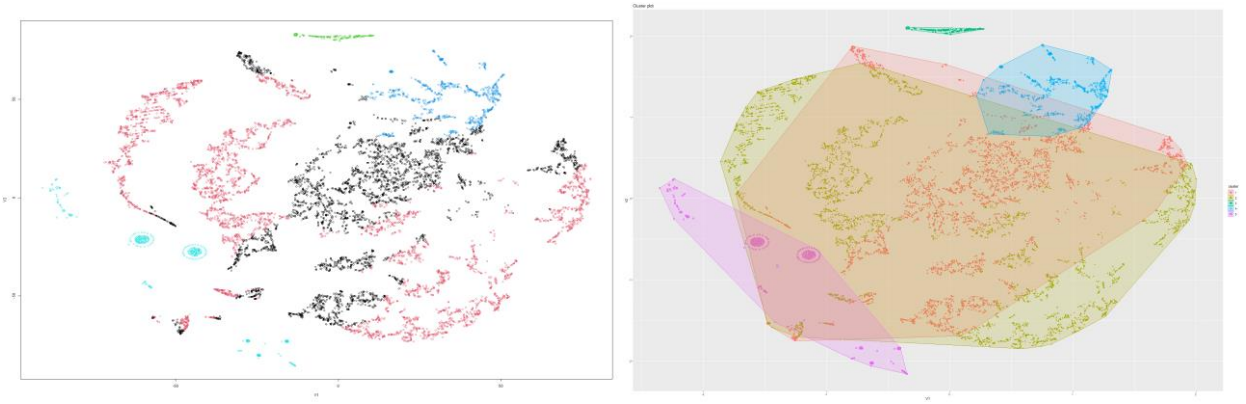


Figure 274&275 k-Means on t-SNE Model 8

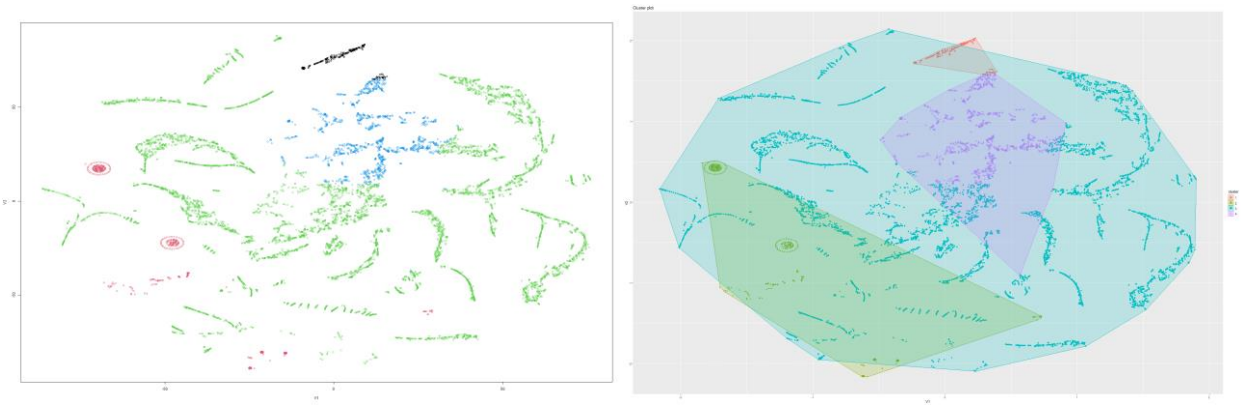


Figure 276&277 k-Means on t-SNE Model 9

6.2.4 K-means on PCA

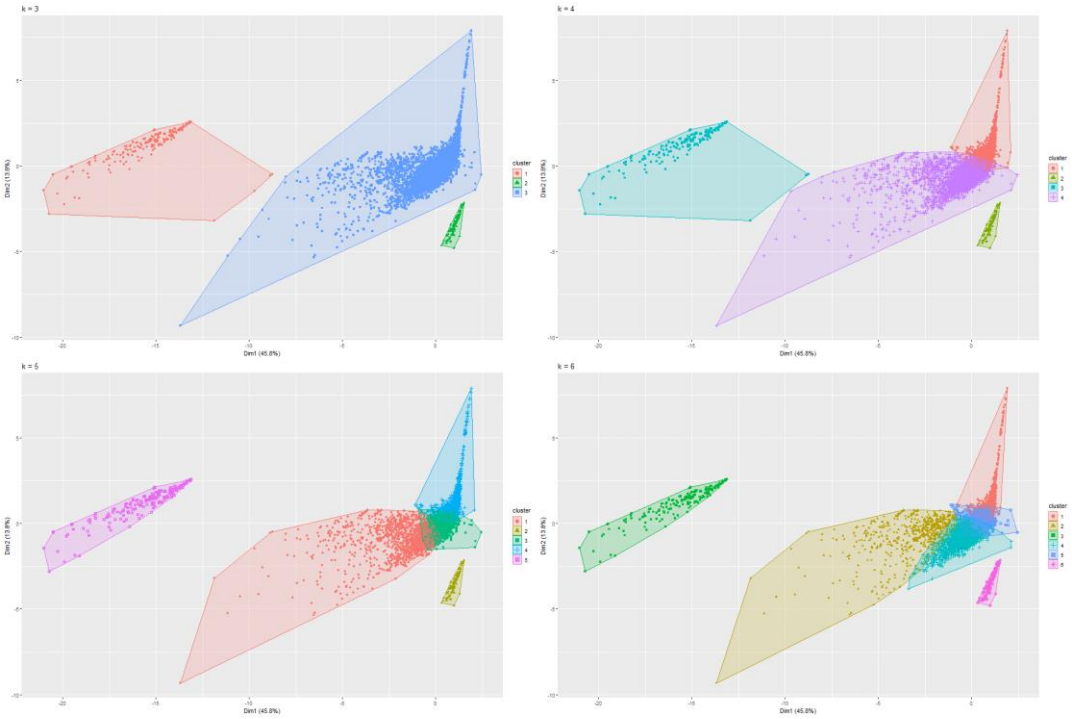


Figure 278 k-Means on PCA Model 1

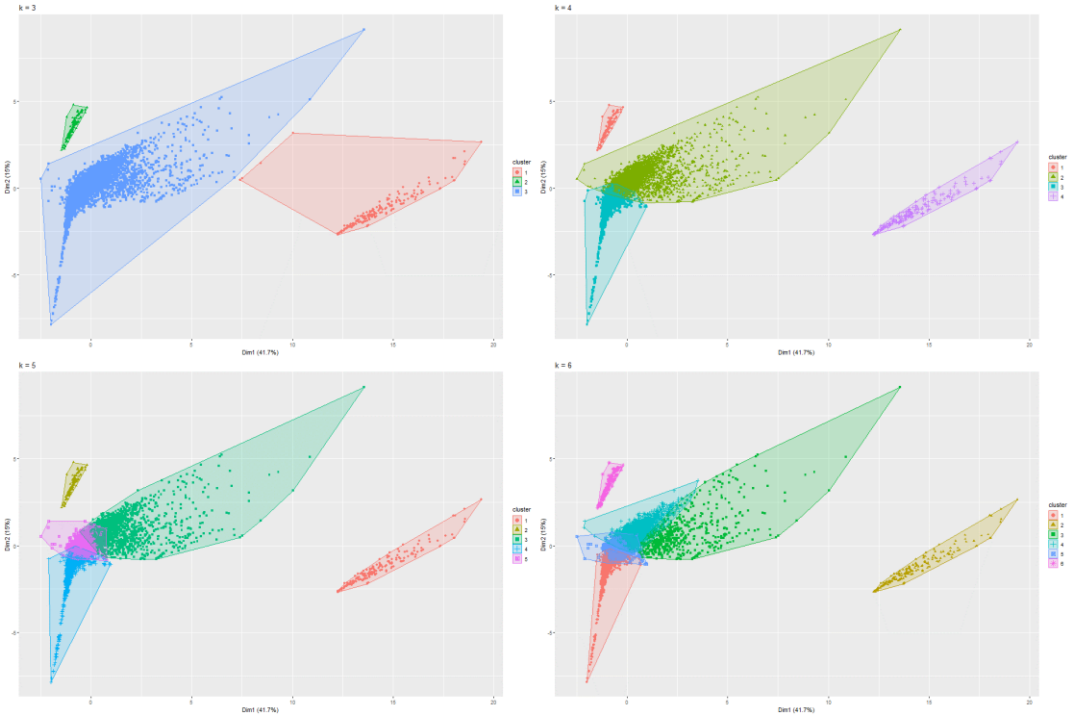


Figure 279 k-Means on PCA Model 2

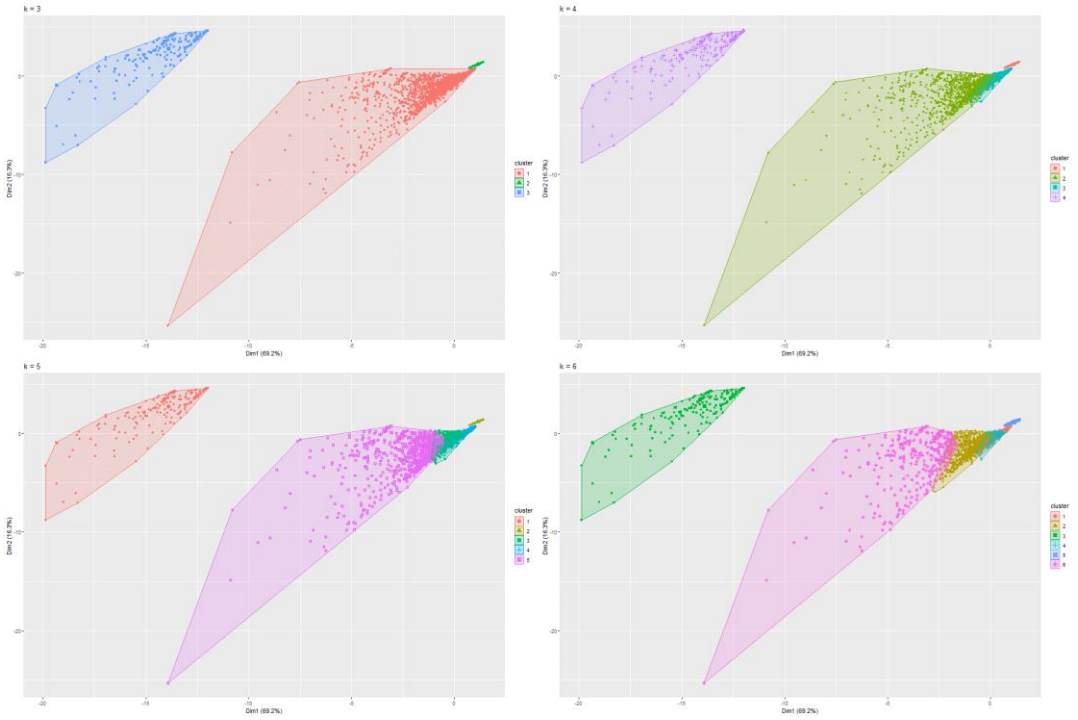


Figure 280 k-Means on PCA Model 3

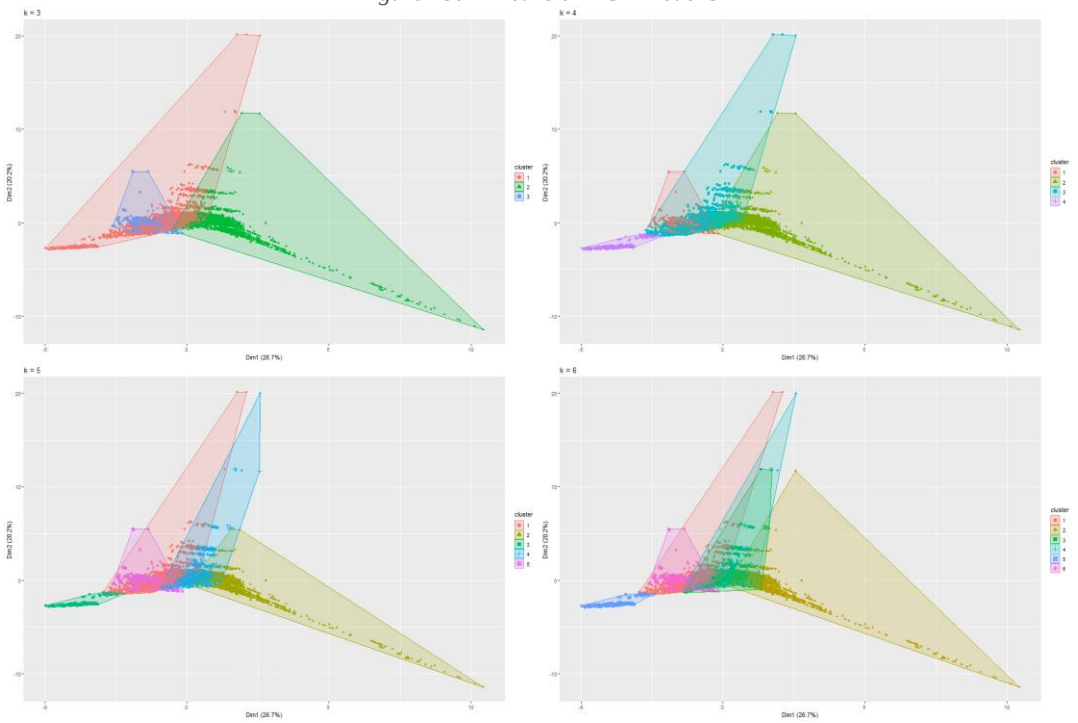


Figure 281 k-Means on PCA Model 4

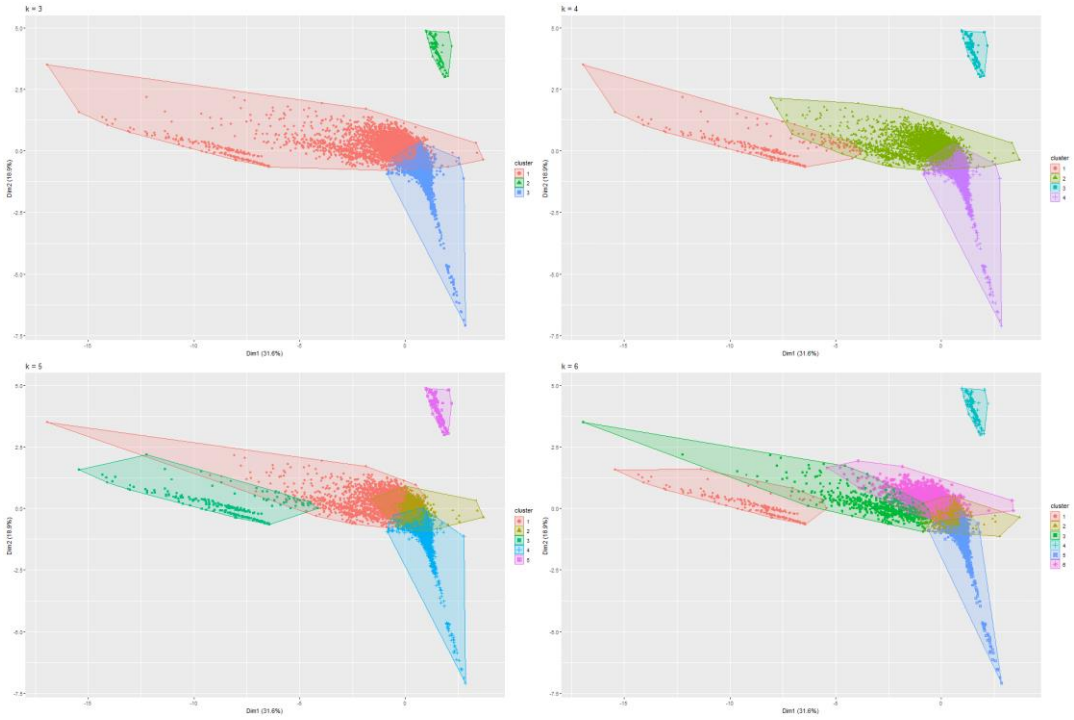


Figure 282 k-Means on PCA Model 5

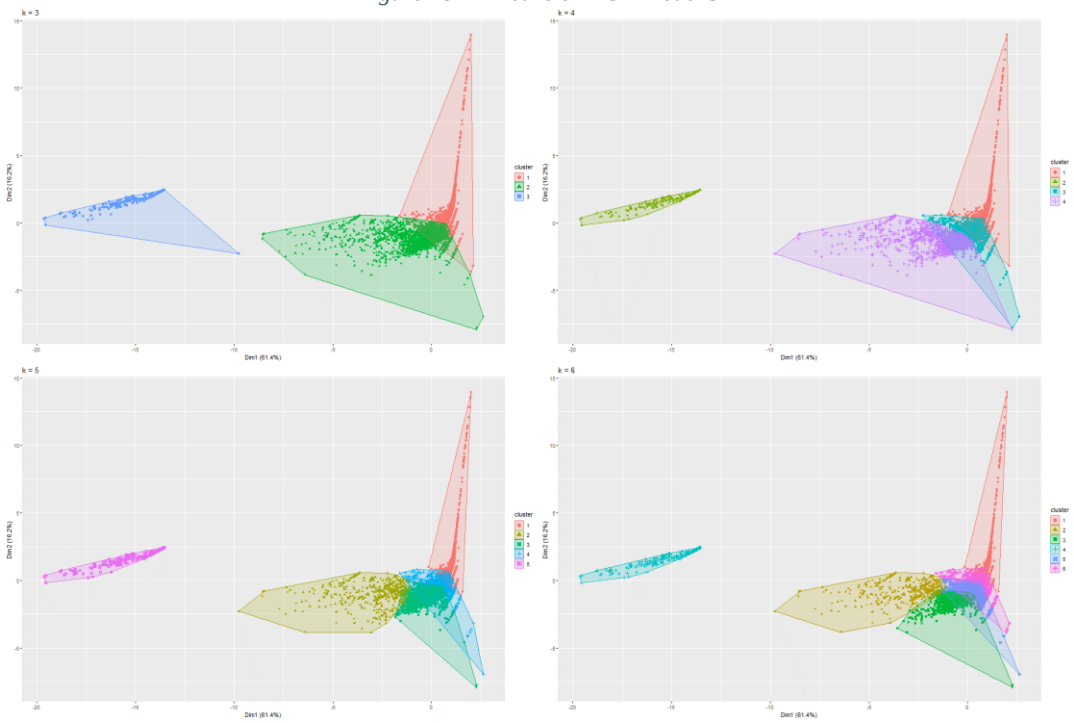


Figure 283 k-Means on PCA Model 6

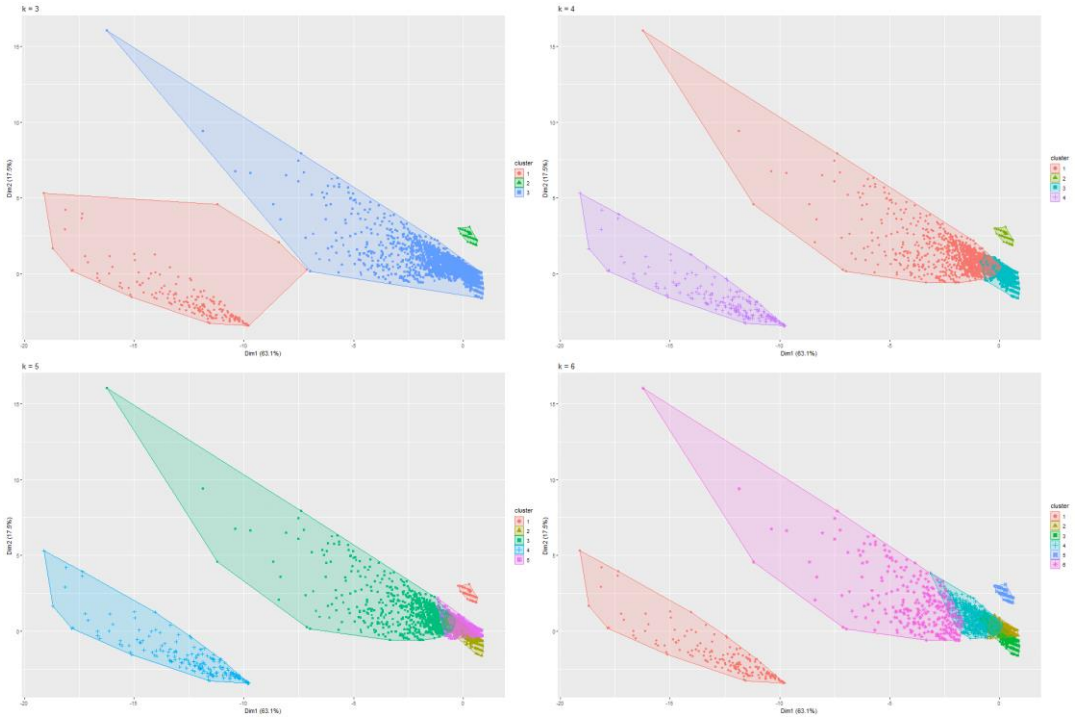


Figure 284 k-Means on PCA Model 7

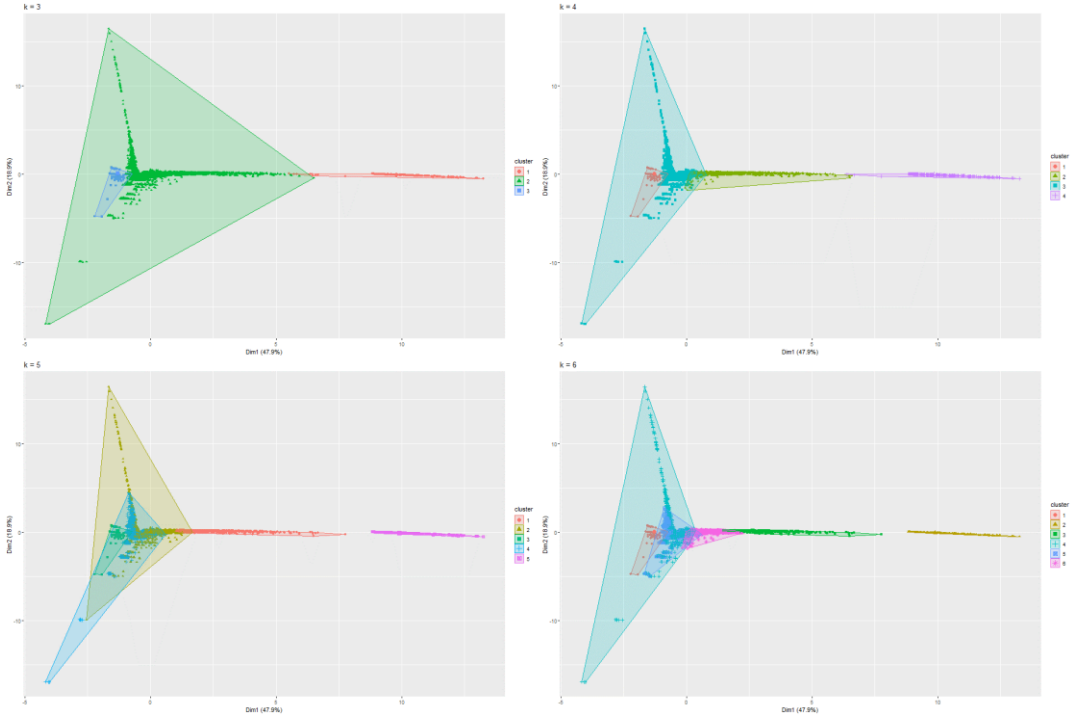


Figure 285 k-Means on PCA Model 8

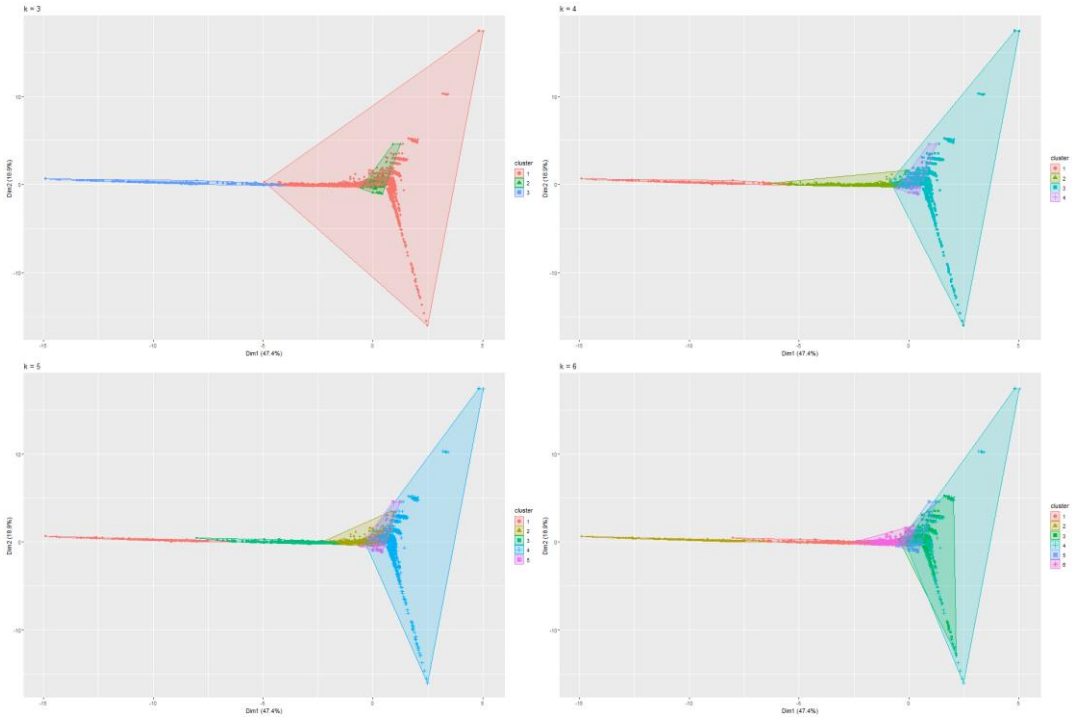


Figure 286 k-Means on PCA Model 9

6.2.5 UMAP and K-means on UMAP

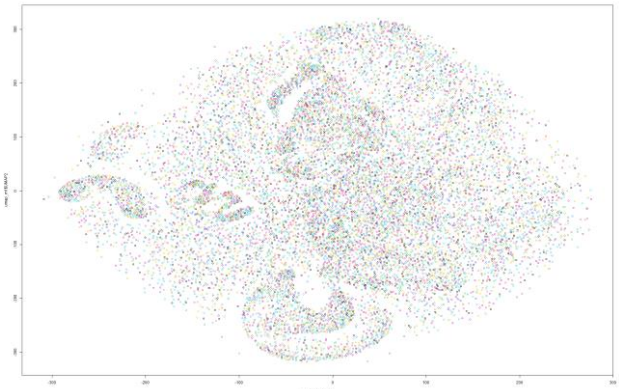


Figure 287 UMAP on Model 1

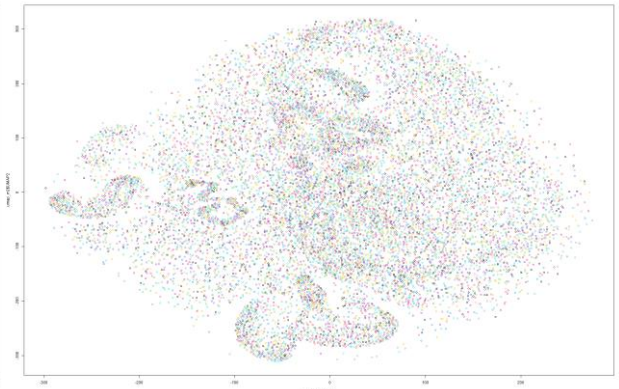


Figure 288 UMAP on Model 2

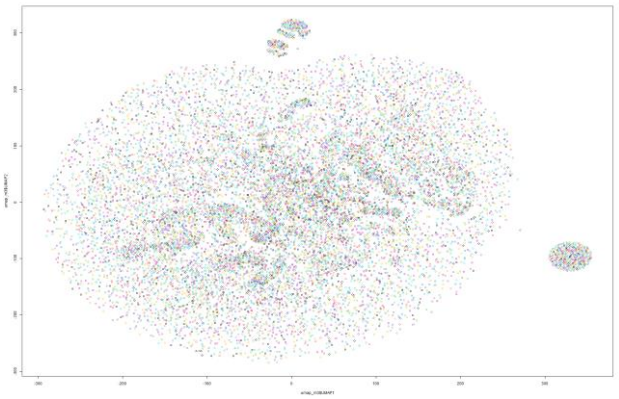


Figure 289 UMAP on Model 3

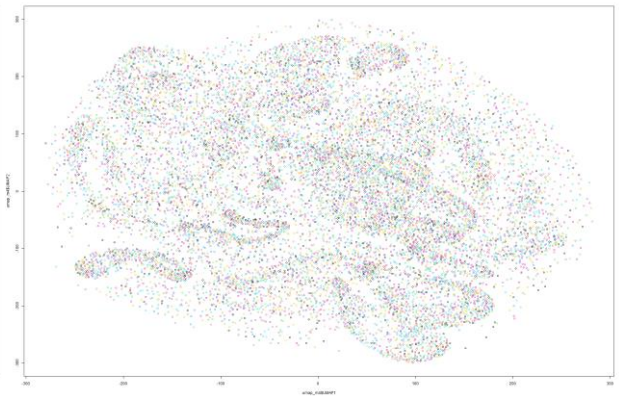


Figure 290 UMAP on Model 4

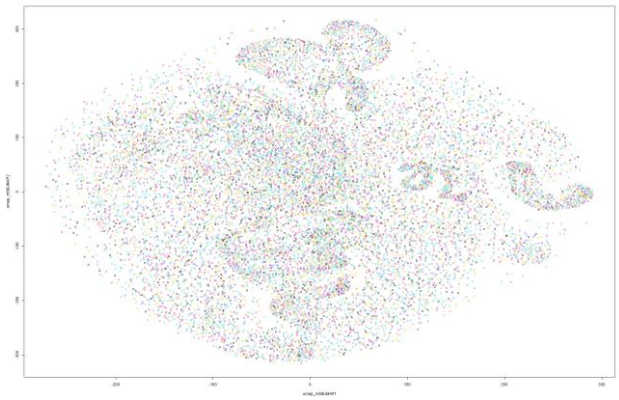


Figure 291 UMAP on Model 5

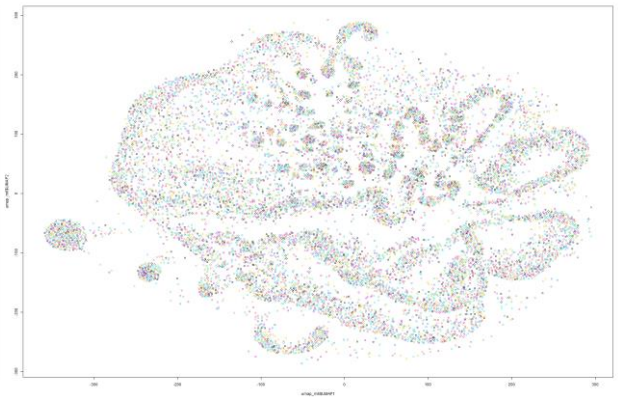


Figure 292 UMAP on Model 6

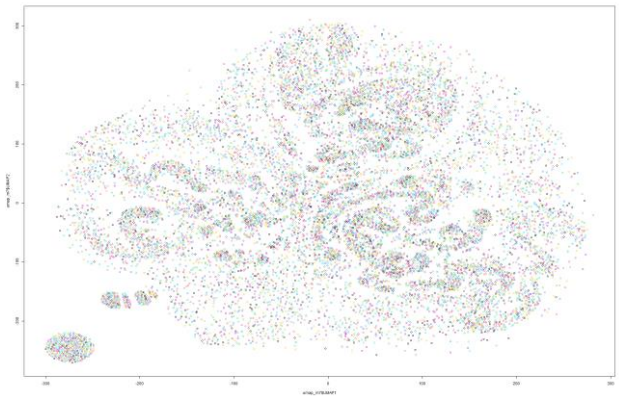


Figure 293 UMAP on Model 7

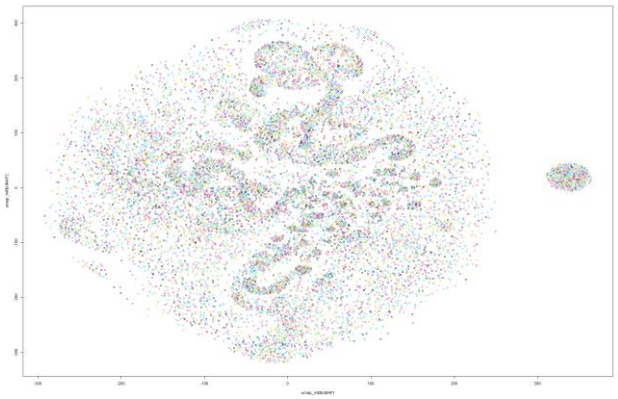


Figure 294 UMAP on Model 8

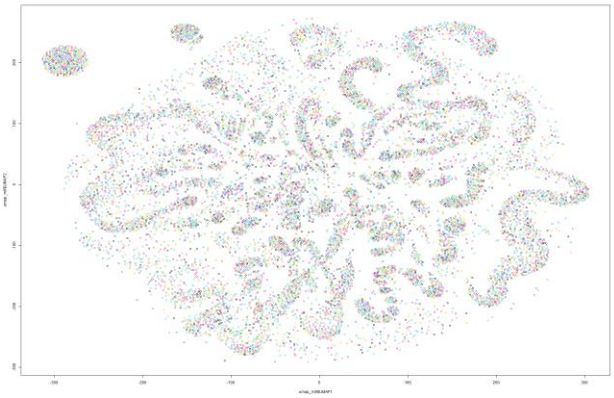


Figure 295 UMAP on Model 9

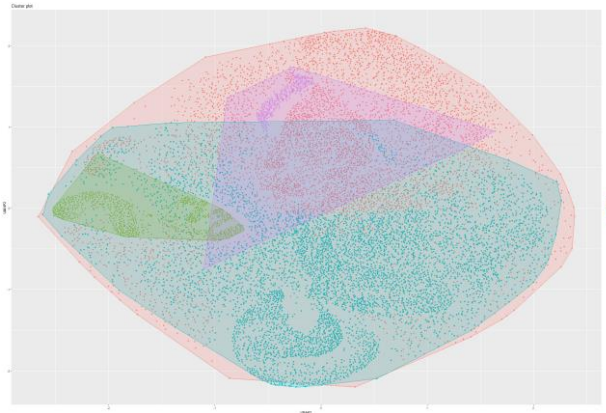


Figure 296 K-means on UMAP on Model 1

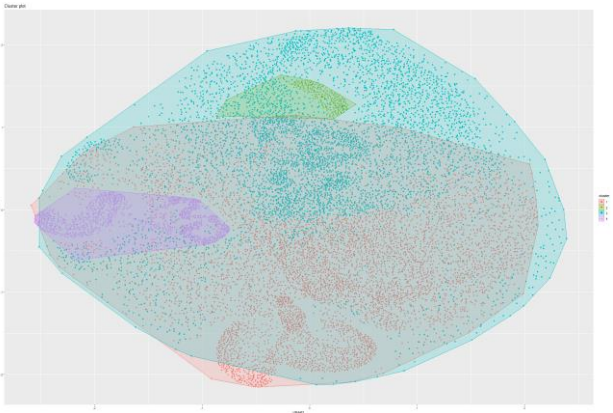


Figure 297 K-means on UMAP on Model 2

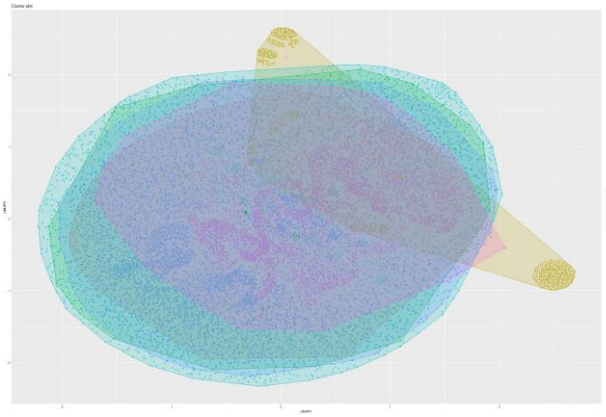


Figure 298 K-means on UMAP on Model 3

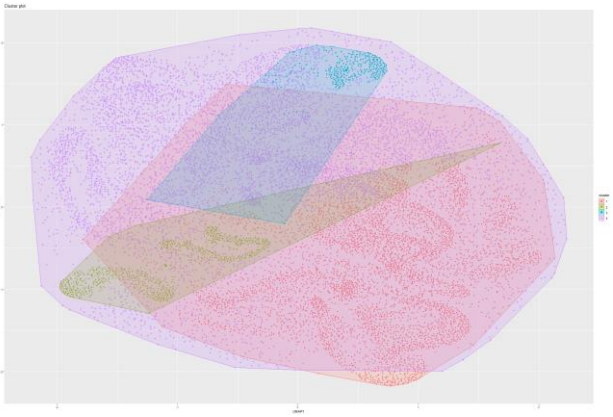


Figure 299 K-means on UMAP on Model 4

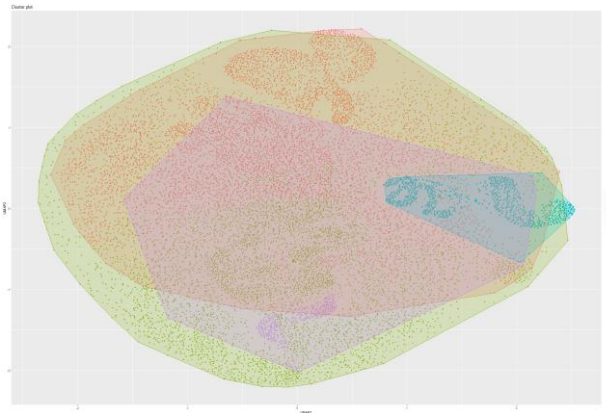


Figure 300 K-means on UMAP on Model 5

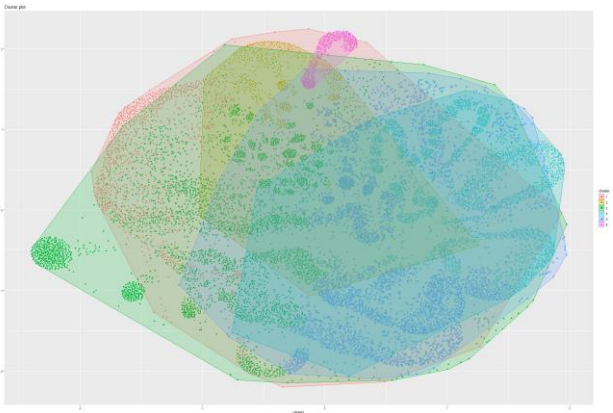


Figure 301 K-means on UMAP on Model 6

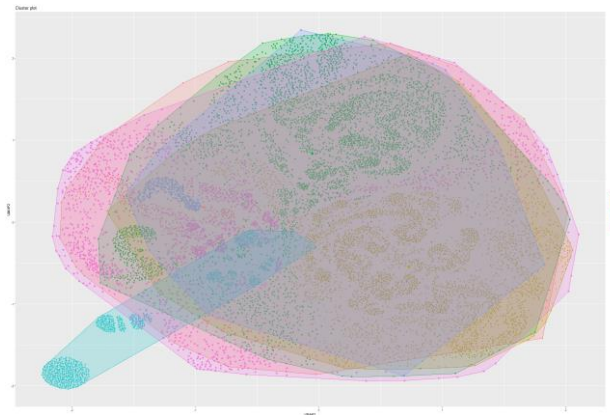


Figure 302 K-means on UMAP on Model 7

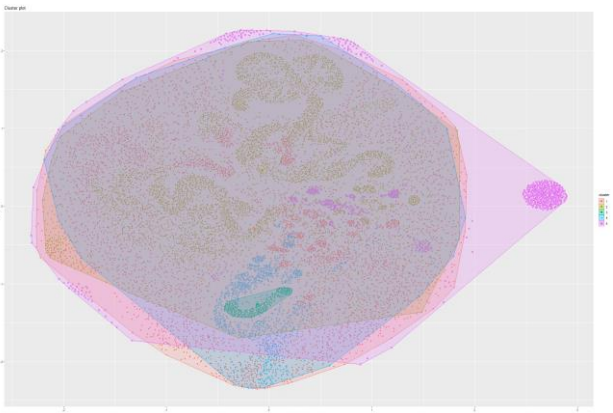


Figure 303 K-means on UMAP on Model 8

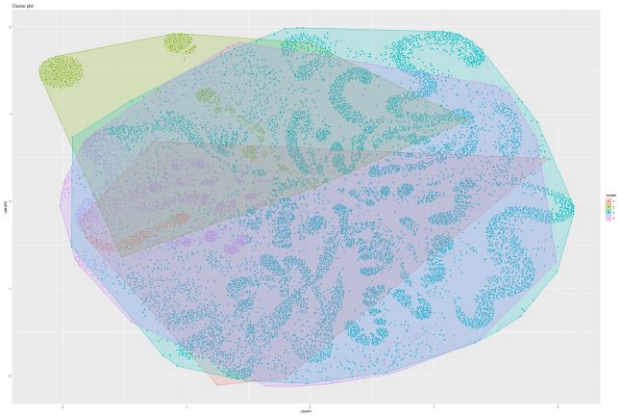


Figure 304 K-means on UMAP on Model 9

6.2.6 Sammon's mapping

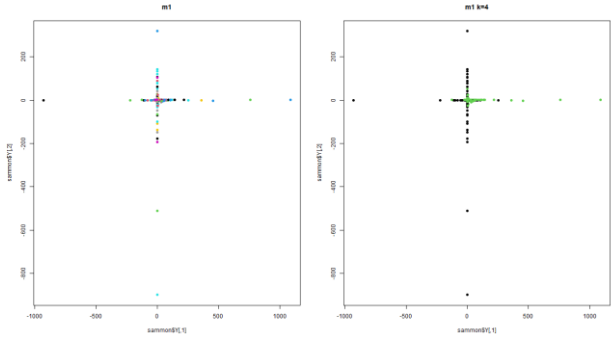


Figure 305 Sammon's mapping on Model 1

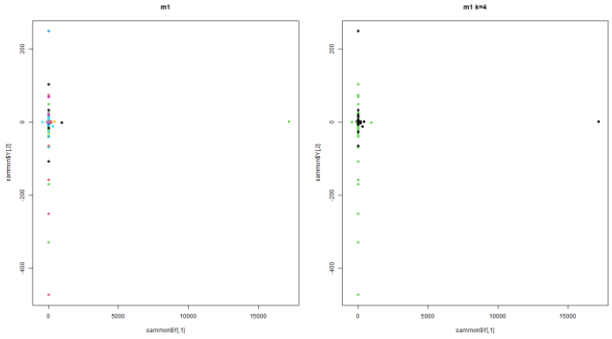


Figure 306 Sammon's mapping on Model 2

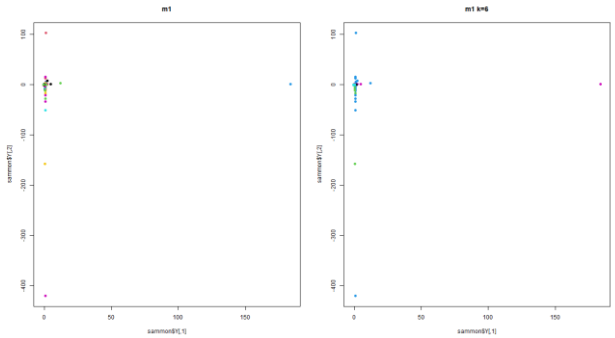


Figure 307 Sammon's mapping on Model 3

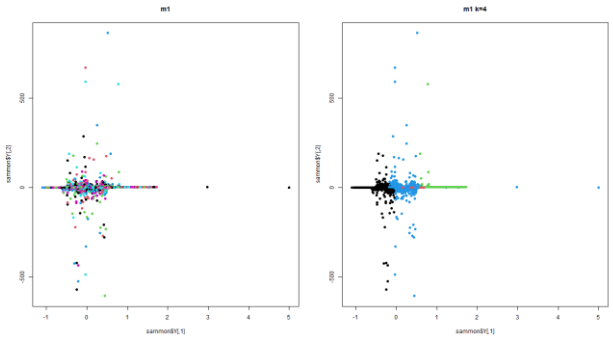


Figure 308 Sammon's mapping on Model 4

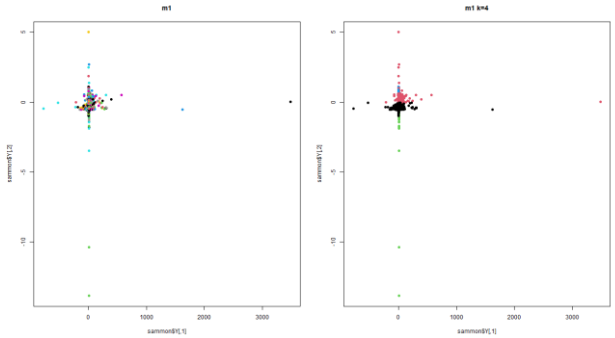


Figure 309 Sammon's mapping on Model 5

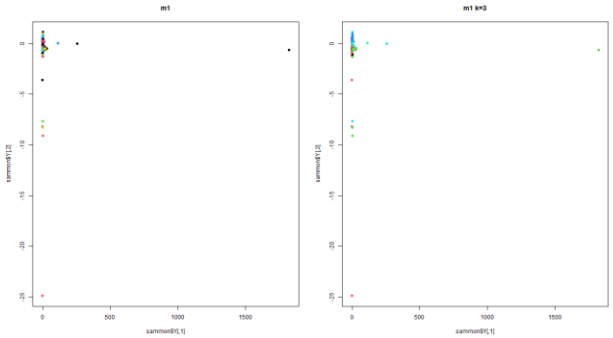


Figure 310 Sammon's mapping on Model 6

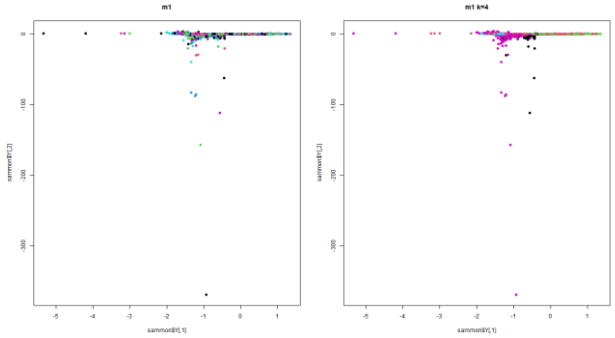


Figure 311 Sammon's mapping on Model 7

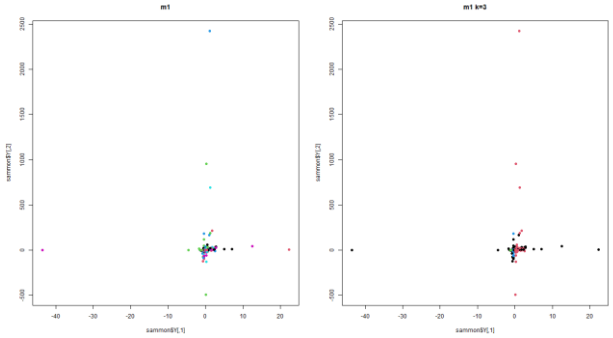


Figure 312 Sammon's mapping on Model 8

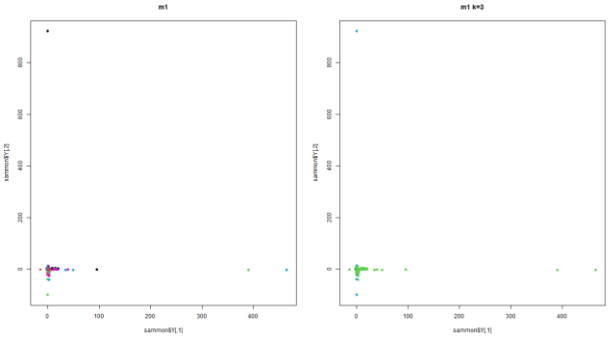


Figure 313 Sammon's mapping on Model 9

6.2.7 DBSCAN

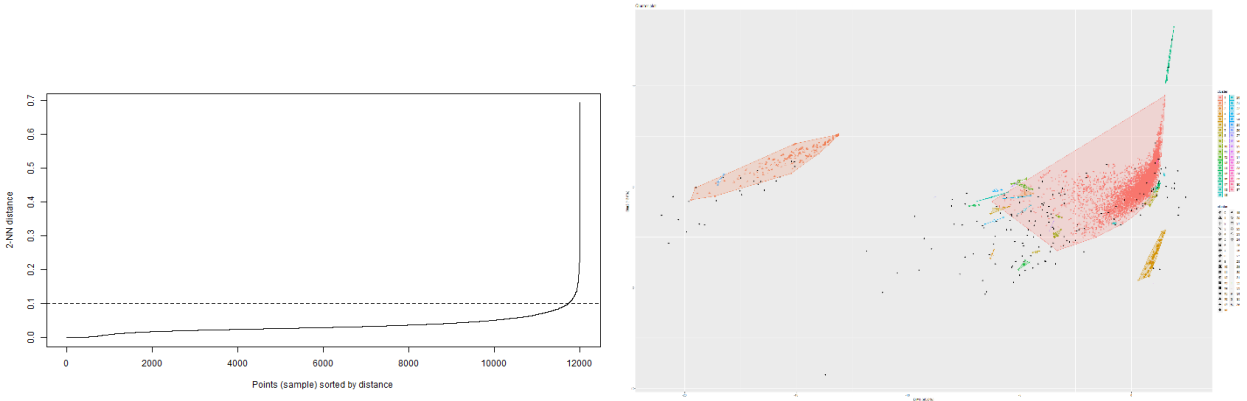


Figure 314 & 315 Knee plot and DBSCAN on Model 1

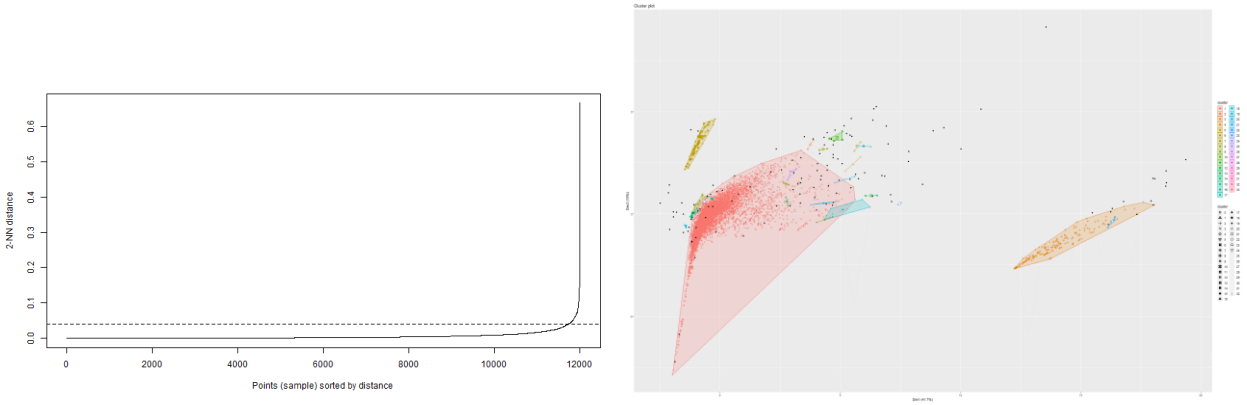


Figure 316 & 317 Knee plot and DBSCAN on Model 2

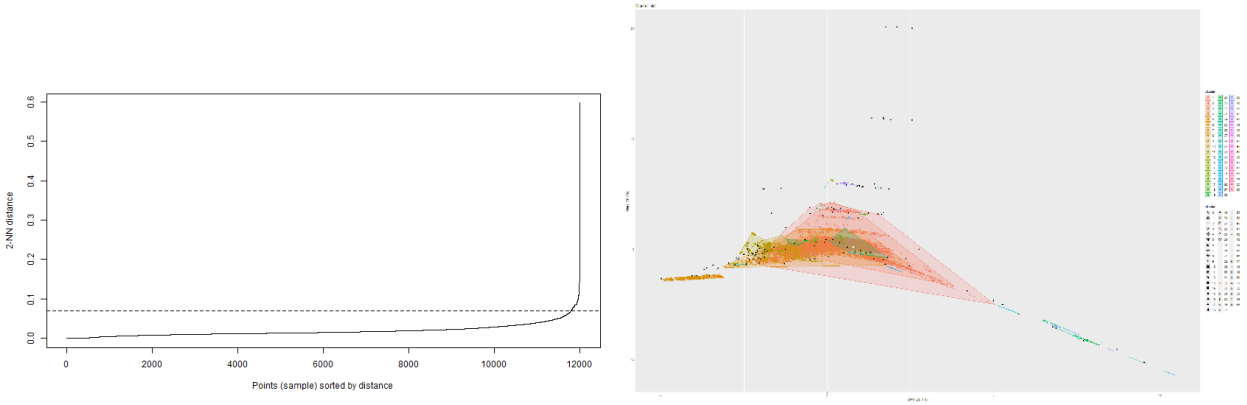


Figure 318 & 319 Knee plot and DBSCAN on Model 4

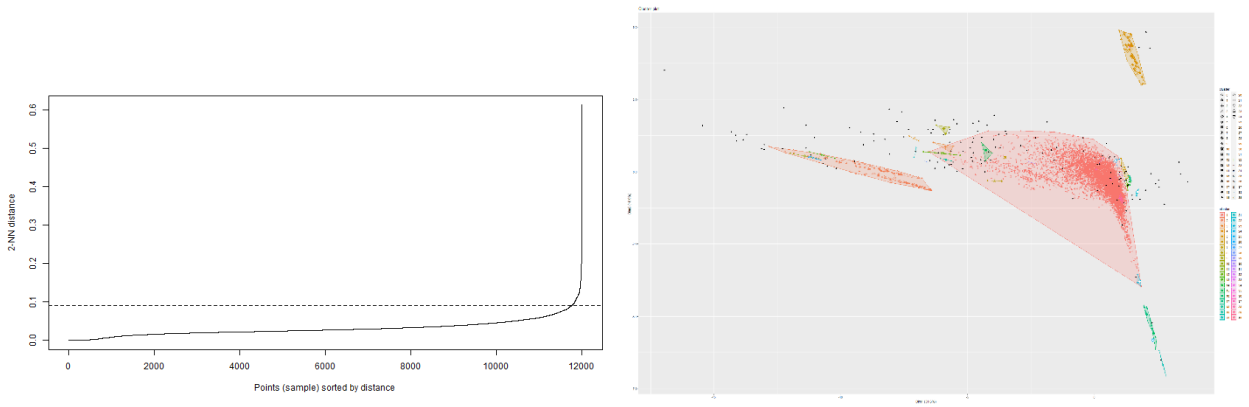


Figure 320 & 321 Knee plot and DBSCAN on Model 5

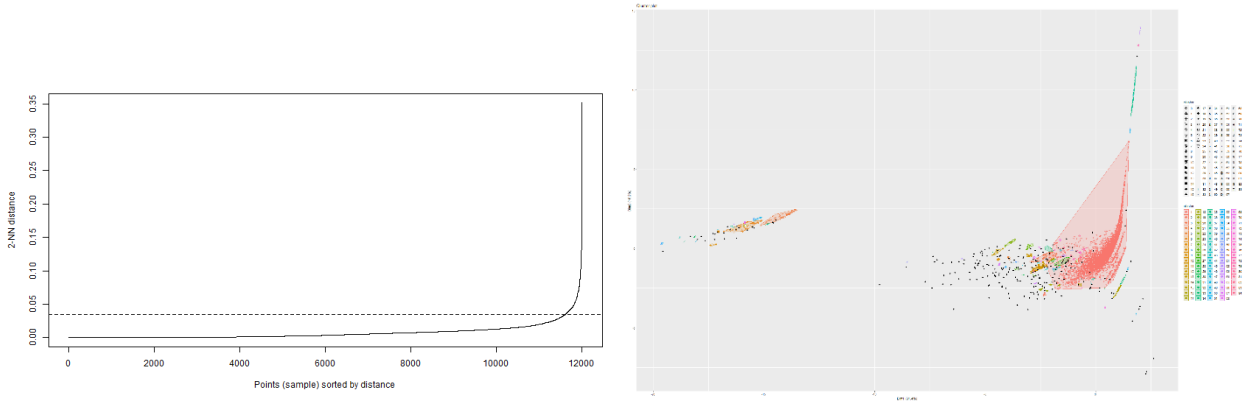


Figure 322 & 323 Knee plot and DBSCAN on Model 6

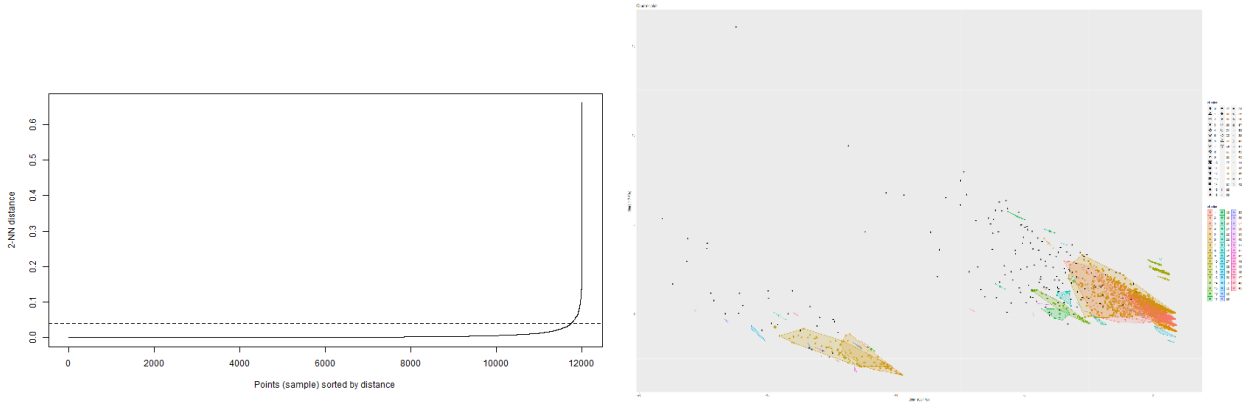


Figure 324 & 325 Knee plot and DBSCAN on Model 7

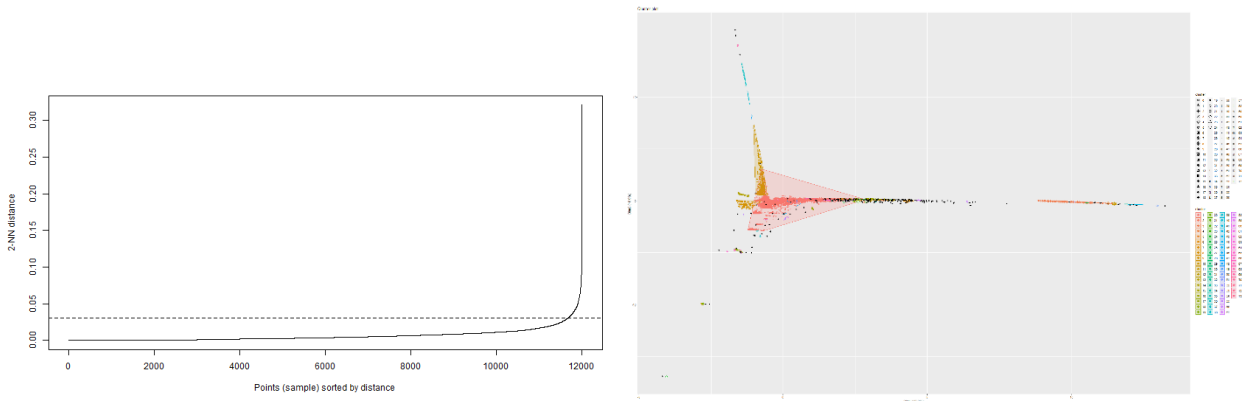


Figure 326 & 327 Knee plot and DBSCAN on Model 8

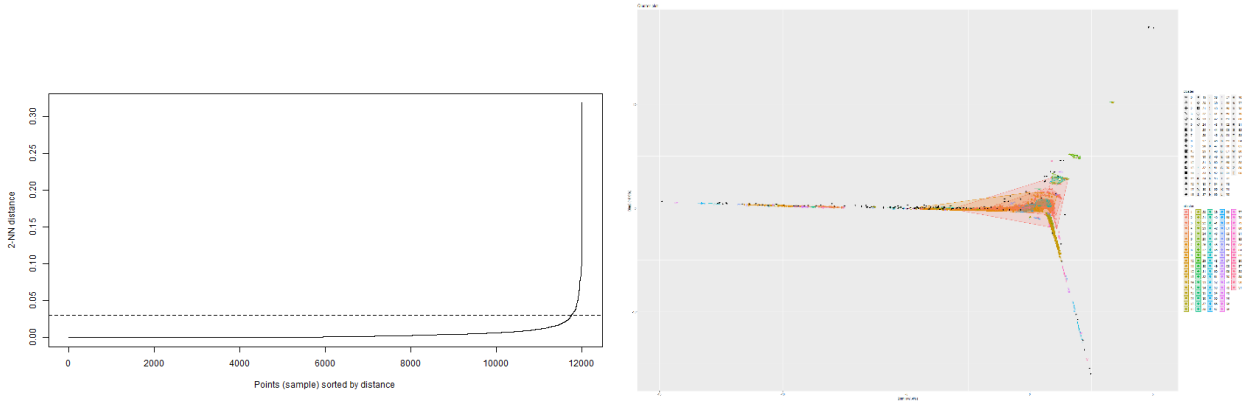


Figure 328 & 329 Knee plot and DBSCAN on Model 9

6.3.1 Discussion

In the sections 6.1.1 and 6.2.1, data analysis was performed on both p2p (Gnutella04) and collaboration network (CA-HepPH). From the boxplot we can see that the variation of values for p2p is much larger than colabn (collaboration network). In the heat map in figure 2 and figure 166 we can also see that p2p has a more prominent positive and negative correlation among its centralities values, whereas colabn has more positive correlation among its centralities. These can be used to discriminate between these networks. However, more descriptive statistics is necessary to be more confident about the discrimination. In the sections 6.1.7 and 6.2.7 dbscan was performed on pca. However, no meaningful cluster formation was not found. Instead, in both networks we can see different cluster groups. Some more dominating than others.

In the sections 6.1.3 and 6.2.3, t-SNE was performed on the datasets. However, afterwards a kmeans clustering was performed on the t-SNE representation. The kmeans fitted a convex hull on the clusters. Optimum number of clusters was derived from the main dataset. The optimum clusters were calculated using knee-plots and Calinski-Harabasz index. Compared p2p, in colabn we can see more meaningful structures. For instance, colabn has circular structures in its clusters. Which can mean that collaborative/ communicative networks are likely to have close clique structures compared to sharing networks like p2p. However, p2p networks do have more abstract structures that are present in all 4 t-SNE models. For instance, in figure 77 and 78 we can see prominent clusters that are not mixed. These can mean community clusters in the network. Same can be said about figure 228 and 229. In those figure we can also see clusters that are not mixed and forming both abstract and simple structures like circles. These patterns can also be seen in the other figures. In the sections 6.1.5 and 6.2.5 umap was performed as an alternative representation compared to t-SNE. However, t-SNE conveys more meaningful representation compared to umap. However, we can see more prominent community formation in colabn models, specifically figures 298, 301, 302, 303 and 304. On the other hand, we can see some form of community structure in figures 132, 135, 136 and 140.

In the sections 6.1.4 and 6.2.4 kmeans was performed on pca. As discussed earlier, optimum cluster numbers were calculated using knee-plots and Calinski-Harabasz index. However, here we fitted different number of clusters apart from the best number. From our ablation models, we can see that both networks have very different spread to their data in dimension 1 and 2 of the pca. Specifically, in model 4 we can see that the data has a stacked structure whereas model 4 of colabn looks more like a camel hump for lack of better words. We can see this similarity on all of the pca results. In the sections 6.1.6 and 6.2.6 sammons map was performed to see the impact of ablation study. Both networks showed discernable impact of the ablations models.

CHAPTER 7: CONCLUSION AND FUTURE WORK

In Conclusion, unsupervised discrimination among networks is possible to some extent. However, more datasets need to be examined and quantified in order to prove that yes different class of networks give different structural output and only than we can easily classify them without having any prior knowledge about them. For future work, we need to able to quantify our lower manifold representation for comparison. If we are able to quantify our visual analysis, we can have a more robust framework for network classification.

REFERENCES

Bell, D. C., Atkinson, J. S., & Carlson, J. W. (1999). Centrality measures for disease transmission networks. Social Networks, 21(1), 1–21. https://doi.org/10.1016/S0378-8733(98)00010-0

Bródka, P., Skibicki, K., Kazienko, P., & Musiał, K. (2011). A degree centrality in multi-layered social network. Proceedings of the 2011 International Conference on Computational Aspects of Social Networks, CASoN'11, 237–242. https://doi.org/10.1109/CASON.2011.6085951

Chen, H., Yin, H., Chen, T., Viet, Q., Nguyen, H., & Xue, W. P. (2019). Exploiting Centrality Information with Graph Convolutions for Network Representation Learning. 2019 IEEE 35th International Conference on Data Engineering (ICDE), 590–601. https://doi.org/10.1109/ICDE.2019.00059

Chiu, C. C., Balkunid, P., & Weinberg, F. (2016). When managers become leaders : The role of manager network centralities , social power , and followers ' perception of leadership. The Leadership Quarterly. https://doi.org/10.1016/j.leaqua.2016.05.004

Cohn, A. M., Amato, M. S., Zhao, K., Wang, X., Cha, S., Pearson, J. L., Papandonatos, G. D., & Graham, A. L. (2019). Discussions of Alcohol Use in an Online Social Network for Smoking Cessation: Analysis of Topics, Sentiment, and Social Network Centrality. Alcoholism: Clinical and Experimental Research, 43(1), 108–114. https://doi.org/10.1111/acer.13906

Cordella, L. P., Foggia, P., Sansone, C., & Vento, M. (2004). A (sub)graph isomorphism algorithm for matching large graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(10), 1367–1372. https://doi.org/10.1109/TPAMI.2004.75

Crucitti, P., Latora, V., & Porta, S. (2006). Centrality in networks of urban streets. Chaos, 16(1). https://doi.org/10.1063/1.2150162

Culotta, A., & Cutler, J. (2016). Predicting Twitter User Demographics using Distant Supervision from Website Traffic Data. 55, 389–408.

Hussain, J., & Islam, M. A. (2016). Evaluation of graph centrality measures for tweet classification. 2016 International Conference on Computing, Electronic and Electrical Engineering, ICE Cube 2016 - Proceedings, 126–137. https://doi.org/10.1109/ICECUBE.2016.7495209

Joyce, K. E., Laurienti, P. J., Burdette, J. H., & Hayasaka, S. (2010). A new measure of centrality for brain networks. PLoS ONE, 5(8). https://doi.org/10.1371/journal.pone.0012200

Landherr, A., Friedl, B., & Heidemann, J. (2010). A Critical Review of Centrality Measures in Social Networks. Business & Information Systems Engineering, 2(6), 371–385. https://doi.org/10.1007/s12599-010-0127-3

Lee, S. H., Choi, J. Y., Yoo, S. H., & Oh, Y. G. (2013). Evaluating spatial centrality for integrated tourism management in rural areas using GIS and network analysis. Tourism Management, 34, 14–24. https://doi.org/10.1016/j.tourman.2012.03.005

Mccallum, A. (2007). Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email. 30, 249–272.

Miller, P. R., Bobkowski, P. S., Maliniak, D., & Rapoport, R. B. (2015). Talking Politics on Facebook : Network Centrality and Political Discussion Practices in Social Media. https://doi.org/10.1177/1065912915580135

Narayanan, S. (2005). The Betweenness Centrality Of Biological Networks A Study of Betweenness Centrality.

Newman, M. (2010). Networks: An Introduction. In Networks: An Introduction. https://doi.org/10.1093/acprof:oso/9780199206650.001.0001

Park, K., & Kim, D. (2009). Localized network centrality and essentiality in the yeast – protein interaction network. 5143–5154. https://doi.org/10.1002/pmic.200900357

Rossman, G., Esparza, N., & Bonacich, P. (2010). I'd like to thank the academy, team spillovers, and network centrality. American Sociological Review, 75(1), 31–51. https://doi.org/10.1177/0003122409359164

Williamson, S. A., & Tec, M. (2019). Random clique covers for graphs with local density and global sparsity. 35th Conference on Uncertainty in Artificial Intelligence, UAI 2019.

Yang, B., & Liu, J. (2008). Discovering global network communities based on local centralities. ACM Transactions on the Web, 2(1). https://doi.org/10.1145/1326561.1326570

Crucitti, P., Latora, V., & Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *73*(3), 1–5. https://doi.org/10.1103/PhysRevE.73.036125

De La Peña Sarracén, G. L., & Rosso, P. (2018). Automatic text summarization based on betweenness centrality. *ACM International Conference Proceeding Series, Part F1377*. https://doi.org/10.1145/3230599.3230611

Department of Physics Department of Electrical and Computer Engineering. (2012). *Review Literature And Arts Of The Americas*, 3977–3980.

Dwyer, T., Kosch, D., & Xu, K. (2003). *Visual Analysis of Network Centralities*. *Wuchty 2002*.

Huang, X., Zhao, Y., Yang, J., Zhang, C., & Ye, X. (2016). *TrajGraph : A Graph-Based Visual Analytics Approach to Studying Urban Network Centralities Using Taxi Trajectory Data*. *22*(1), 160–169.

Wu, Q., Qi, X., Fuller, E., & Zhang, C. Q. (2013). "follow the leader": A centrality guided clustering and its application to social network analysis. *The Scientific World Journal*, *2013*. https://doi.org/10.1155/2013/368568

Zhang, Y., Wang, X., Zeng, P., & Chen, X. (2011). Centrality characteristics of road network patterns of traffic analysis zones. *Transportation Research Record*, *2256*, 16–24. https://doi.org/10.3141/2256-03