이학석사 학위논문

A method getting the transitive kNN graph of a kNN graph preserving topology

(kNN 그래프로부터 위상을 보존하는 transitive kNN 그래프를 얻는 방법)

2020 년 8 월

서울대학교 대학원
수리과학부
박정호

A method getting the transitive kNN graph of a kNN graph preserving topology

지도교수 강 명 주

이 논문을 이학석사 학위논문으로 제출함

2020 년 6 월

서울대학교 대학원
수리과학부
박 정 호

박정호의 이학석사 학위논문을 인준함

2020 년 6 월

위 원 장 ________________ (인)
부 위 원 장 ________________ (인)
위 원 ________________ (인)

A method getting the transitive kNN graph of a kNN graph preserving topology

by

Jung Ho Park

A THESIS

Submitted to the faculty of the Graduate School
in partial fulfillment of the requirements
for the degree Master of Science
in the Department of Mathematical Sciences
Seoul National University
August 2020

Abstract

A method getting the transitive kNN graph of a kNN graph preserving topology

Jung Ho Park
Department of Mathematical Sciences
The Graduate School
Seoul National University

In this paper, we recall some previous research of a bijection between transitive directed graphs and topologies on a finite set. In data science area, k Nearest Neighbor(kNN) graph is widely used graph. It can be interpreted by a directed graph in the sense of in and out degrees. kNN graph is not transitive in general. Adding or eliminating edges, there are various methods to get a transitive directed graph from a given kNN graph. In this paper, we prove that there is a unique way which make a transitive directed graph preserving topology. Moreover, we suggest a simple method to make the transitive directed graph without computing topology of a given kNN graph.

Keywords : kNN graph, transitive directed graph, finite topology
Student number : 2017–28387

Contents

1 Introduction

In almost previous researches [2, 3, 4], the relation between finite topologies and transitive directed graphs is used to compute the number of distinct topologies on a finite set. However, many application areas based on data analysis want to extract information of graphs which comes from datasets. For instance, [1] interprets dataset by embedding a topological manifold and uses graph theory and Morse theory to explore the dataset. This paper is my first step to analyze graphs by a mathematical aspect via topology. kNN graph is a graph widely used in many application areas. Briefly, in a metric space, kNN graph contains information of points and k-th closest edges for each point. In partiacular, we can give a direction to kNN graph by in-degree and out-degree. kNN directed graph may be properly used to datasets whose features have a partial ordered structure. In general, a directed graph is neither transitive nor acyclic so that it makes troubles on applications. In this paper, assuming given directed graphs are acyclic, we concentrate on treatment of transitivity of directed graphs. In section 2.4, we can give mappings between directed graphs and topologies on a finite set. We may think that the it is important edge whose end-points are in a open set. In this sense, for a given kNN graph, we extract subgraph consisted of important edges in the sense above. It is to be a transitive directed subgraph preserving the corresponding topology. In the section 3, we suggest theorems 3.1 and 3.2 which give a explicit method getting a transitive directed graph of given directed graph. In section , we show directed kNN graphs has their transitive directed kNN subgraph, and introduce an algorithm as an application.

2 Preliminaries

In this section, we briefly review backgrounds of this paper; some notions of graph theory, topology, and kNN graph. And then, we introduce a bijection between the collection of topologies and the collection of transitive directed graph on a finite set.

2.1 Notions of finite directed graphs

Graph is a useful tool when ones explore complicated manifolds or represent relation of elements in a dataset. In paticular, directed graph sometimes use to make a model about flow of information. First, let us remind following basic definitions.

Definition 2.1. For a finite X, a tuple (X, G) is called a directed graph if G is a subset of $X \times X - \{(x, x) : x \in X\}$. We call elements in X, G as vertices, edges, respectively. We, sometimes, denote edge (x, y) as xy.

Definition 2.2. For an edge (x, y) in a finite directed graph (X, G), we say x is adjacent to y.

Definition 2.3. For a directed graph (X, G) and a finite sequence $\{x_i\}_{i=1}^n$ in X such that $x_i x_{i+1} \in G$, we say $\{x_i\}_{i=1}^n$ is a cycle if $x_1 = x_n$.

Definition 2.4. A finite directed graph (X, G) is transitive if, for any $x, y, z \in X$, both (x, y) and (y, z) are in G implies so (x, z) is.

Definition 2.5. In this paper, we deonte $\mathcal{G}_X$ is the collection of directed graphs on X. Moreover, $\mathcal{G}_X^T$ means the collection of transitive directed graphs on X.

Definition 2.6. A finite directed graph (X, G) is asymmetric if, for any $x, y \in X$, $(x, y) \in G$ implies $(y, x) \in G$.

Definition 2.7. A directed graph is called acyclic if the graph has no any cycle.

Definition 2.8. For a given directed graph $G = (X, V)$, we define two maps $\text{indeg} : X \to \mathbb{Z}$ by

$$\text{indeg}(x) \quad \text{is the number of edges whose end point is } x \quad x \in X \tag{2.1}$$

and $\text{outdeg} : X \to \mathbb{Z}$ by

$$\text{outdeg}(x) \quad \text{is the number of edges whose start point is } x \quad x \in X. \tag{2.2}$$

2.2 Adjacency matrices

Representing matrix of a directed graph has many merits. One of the most merits is that there is no loss of information of graph. i.e., we can perfectly recover the directed graph from the matrix representation. Moreover, since the set of matrices forms a ring structure, it is more flexible way to analyze directed graphs. We call the represented matrix as the adjacency matrix. In this section, we see the definition of adjacency matrix and its application.

Definition 2.9. Let $X = \{1, 2, \cdots, N\}$ for an integer N and (X, G) be a directed graph. A matrix $A(G)$ in $\mathcal{M}_{N\times N}(\mathbb{Z}_2)$ is the adjacency matrix of G which is defined by

$$A(G) = (a_{ij}) \quad \text{where} \quad a_{ij} = \begin{cases} 1 & ,(i,j) \in G \\ 0 & ,\text{otherwise} \end{cases}. \tag{2.3}$$

Lemma 2.10. *Let us assume $f : \mathcal{G}_X \to \mathcal{M}_{N\times N}(\mathbb{Z}_2)$ which is defined by the definition 2.9. Then, f is bijective.*

Definition 2.11. Let $X = \{1, 2, \cdots, N\}$ and G be a directed graph on X. The in-degree of j-th point is the number of the edges whose end point is j. Similarly, the out-degree of the point i is the number of edges with start point is i.

observation 1. For a directed graph G on $\{1, 2, \cdots, n\}$, consider the adjacency matrix $A(G)$. Let us denote

$$A(G) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \ddots & \ddots & & \ddots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Then, for a fixed $j \in \{1, 2, \cdots, n\}$, the in-degree of the j-th point is the sum of elements of j-th column of $A(G)$, i.e., $\mathrm{indeg}(j) = \sum_{i=1}^{n} a_{ij}$. By the same way, for fixed $i \in \{1, 2, \cdots, n\}$, the out-degree $\mathrm{outdeg}(j)$ of the i-th point is $\sum_{j=1}^{n} a_{ij}$.

observation 2. Path is a finite sequence of points. We denote a path with length L as $\{x_i\}_{i=1}^{L}$ where $L \in \mathbb{N}$ and $x_i \in G$. For a given directed graph G, we can easily get the matrix which is consisted of the number of paths with length L by the adjacency matrix A of G by

$$A^L \in \mathcal{M}_{n\times n}(\mathbb{Z}). \tag{2.4}$$

2.3 Notions of finite topological spaces

Topology is a kind of area in mathematics which study geometric objects. To treat them, it is necessary to distinguish two points and to express closeness. Explicit distance of two points in object does not matter in topology. In topology, open sets play key role, since those give these information. Now let us remind basic definitions in topology.

Definition 2.12. A set X, a topology T on X is a subcollection of $\mathcal{P}(X)$ which contains $\emptyset$ and X ,and is closed under finite intersection and any union. Every elements in a topology is called an open set, and the complement of an open set is called a closed set.

Example 1. One of the most simple topologies on X is $\mathcal{P}(X)$ which is called a discrete topology.

Definition 2.13. In this paper, we denote $\mathcal{T}_X$ as the collection of topologies on X.

Definition 2.14. A basis $\beta \subset T$ of a topology T is an open cover of X with following property: for $B_1, B_2 \in \beta$, and for all $a \in B_1 \cap B_2$, there is B_3 such that $x \in B_3 \subset B_1 \cap B_2$.

It is important the domain set X is whether discrete or continuous. In this paper, we treat even finite set. Over such domain X, 'finite' intersection is the same to 'any' intersection.

Definition 2.15. A topological space is called T_0 space when, for distinguishable two points, there is an open neighborhood of one of points which does not contain the others.

Definition 2.16. A topological space is called T_1, when every singleton set is closed.

Lemma 2.17. *On finite topology, every T_1 topology is a discrete topology.*

By this lemma 2.17, if more strong condition of T_0 make the topology discrete. Thus, we observe only T_0 spaces on a finite set.

2.4 A Bijection between transitive directed graph and topology

In this section, we introduce a bijection between topologies and transitive directed graphs on a finite set denoted by $\Phi : \mathcal{T}_X \to \mathcal{G}_X^T$ and its inverse Ψ. Let T be a topology on a finite set X. Put O_x as the smallest open neighborhood containing x in T. Then, we can get a directed graph getting together adjacent points of x as elements in $O_x \setminus \{x\}$. For a convenience, let us put two notions for a directed graph (X, G).

$$x\downarrow := \{x\} \cup \{y : (y, x) \in G\} \quad \text{and} \quad G\downarrow = \{x\downarrow : x \in X\}. \tag{2.5}$$

i.e., $x\downarrow$ is the set consisting x and all its adjacent points. We can derive the topology T generated by $G\downarrow$ as an open basis. Since X is a finite set, T is generated by finite union and intersection of elements in $G\downarrow$. This is the explicit map Ψ.

Lemma 2.18. *For a finite set X, a map Φ sends topologies to directed graphs. For any topology T, explicitly, $(x, y) \in \Phi(T)$ satisfies that distict $x, y \in X$, x is in every open neighborhood of y. Then, Φ is injective and $\Phi(T)$ is a transitive directed graph.*

Proof. First show the well-definedness of $\Phi : \mathcal{T}_X \to \mathcal{G}_X$. Suppose that G, G' are distinct directed graph which come from T. Since $G \neq G'$, there are distinct $x, y \in X$ such that (x, y) is in either G or G'. Without loss of generality, assume (x, y) is in G but not in G'. By the definition of Φ, (x, y) is in G means that x is in every open neighborhood of y. However, this contradicts to $(x, y) \notin G'$. Hence, Φ is well-defined. In order to show injectivity, arbitrary take different topologies T, T' on X. Then, there is $x \in X$ such that

$$\bigcap_{x \in U \in T} U \neq \bigcap_{x \in U' \in T'} U'. \tag{2.6}$$

No loss of generality, we may take y such that $y \in \bigcap_{x\in U\in T} U$ and $y \notin \bigcap_{x\in U'\in T'} U'$. Then, $(y,x) \in \Phi(T)$ and $(y,x) \notin \Phi(T')$ so that Φ is injective. Let's show $\Phi(T)$ is transitive. For distinct $x,y,z \in X$, assume that x is adjacent to y, y is adjacent to z. For any open neighborhood U of z, y is in U. Since U is an open neighborhood of y and x is adjacent to y, x is in U, too. Hence, x is adjacent to z so that $\Phi(T)$ is transitive. □

Lemma 2.19. *Define a map $\Psi : \mathcal{G}_X \to \mathcal{T}_X$ by $\Psi(G)$ is a topology with a basis $G\downarrow$, where $G\downarrow = \{x\downarrow : x \in X\}$. Then, Ψ is surjective and $\Psi \circ \Phi = id_{\mathcal{T}_X}$, but $\Phi \circ \Psi \neq id_{\mathcal{G}_X}$,in general.*

Theorem 2.20. *Restricting the range of Φ from $\mathcal{G}_X$ to $\mathcal{G}_X^T$, Φ is a bijection. Moreover, $\Phi \circ \Psi = id_{\mathcal{G}_X^T}$.*

Proof. By the lemma 2.18, we know Φ is a well-defined injective map. Let's show surjectivity of $\Phi : \mathcal{T}_X \to \mathcal{G}_X^T$. For any transitive directed graph G, let's take a topology $\Phi^{-1}(G) = \Psi(G)$. Now the rest is to check $\Phi(\Phi^{-1}(G)) = G$. By transitivity of G, $x\downarrow$ is the smallest open neighborhood of x in $\Phi^{-1}(G)$. Therefore, Φ is a bijection. □

Theorem 2.21. *Let $(X,G) \in \mathcal{G}_X$ be a cycle and $\Psi(X,G) = (X,T)$. Then, T is the discrete topology.*

2.5 Local Maximum Feature selection Algorithm

In previous study [1], we introduced a feature selection method called **Local Maximum Feature selection Algorithm**. It is a method to select independent features in a table type dataset. We use kNN graph in this algorithm. Thus, we breifly introduce our previous study as an example of usage of kNN graph. One novelty of this study is that we embed a dataset into a real projective space. We can give a metric on a real projective space space so that we make a kNN graph of dataset. The following is the steps of our algorithm:

Algorithm 1: Local maximum feature selection algorithm

Input: Dataset with features
Output: Selected features

1. embed dataset to $\mathbb{R}P^{N-2}$, where N is the size of dataset
2. make a kNN graph by Hausdorff distance
3. compute normalized Morse index by a score function.
4. take features with local maximum with parameter radius r

For diagnosis diseases, many medical datasets are table type. Such datasets consist of many features of conclusions of medical checks. Hence, let us think a table type medical dataset with N-th features. In the point of view of machine learning, we select independent features, or medical checks, to improve the accuracy of diagnosis. Features presented by different units, so we normalize each features first. Normalized data can be embbeded into a real projective space with the Housdorff metric, and make a kNN graph. Fisher score is well-known to measuring independency of dataset. To select features with local maximum of Fisher score, we use the normalized Morse index.

Definition 2.22. $(N-2)$ dimensional projective space $\mathbb{R}P^{N-2}$ is a topological manifold defined by

$$S^{N-2}/(x \sim -x), \tag{2.7}$$

where S^{N-2} is the unit sphere in $\mathbb{R}^{N-1}$. Furthermore, we can give a metric ρ on $\mathbb{R}P^{N-2}$ by the Housdorff distance which formulated by, for $x, y \in \mathbb{R}P^{N-2}$,

$$\rho(x,y) = \min\{\angle(x,y), \pi - \angle(x,y)\} = \arccos|corr(x,y)|. \tag{2.8}$$

where $\angle(x,y)$ is the geodesic distance on S^{N-2}.

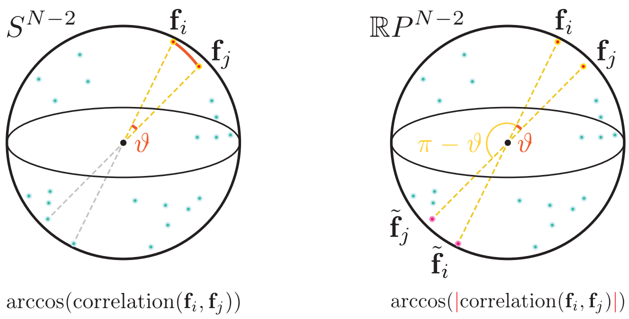


Figure 1: Embedding a dataset on S^{N-2} and $\mathbb{R}P^{N-2}$ in the step 1 in the algorithm 1

Definition 2.23. For a non-empty discrete set X with metric ρ, a k Nearest Neighbor (kNN) graph (X, G) is defined by

$$G = \{(x, a) : \forall x \in X, a \text{ is one among nearest } k \text{ points in } (X, \rho)\}. \tag{2.9}$$

Definition 2.24. The normalized Morse index(NMI) is a function defined on the vertices set X of (X, G) by

$$\text{NMI}(x) = \frac{\#(\text{neighbor whose score is lower than } x)}{\text{degree of } x}. \tag{2.10}$$

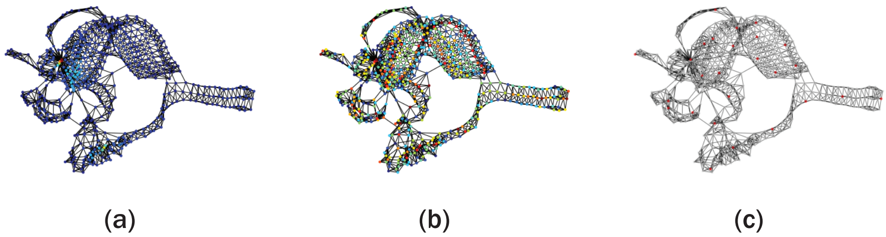


Figure 2: The red point in (a) is the global maximum of score function, the red points in (b) are local maximum of score function in the step 4, and (c) is the conclusion of feature selection in the algorithm 1

3 Algorithm getting a transitive directed graph from a directed graph

In [2, 3, 4], the relation of finite topologies and transitive directed graphs is only used to compute the number of different topologies on X. From a basis, we may naturally get a topology. Observing the proof of lemma 2.20, we may know the open balls with radius 1 forms the basis exactly. In T_0 topological space, we can not distinguish two points in the same open ball with radius 1. Therefore, we may interpret those are closed in the sense of topology. However, the lemma 2.20 is the statement only for transitive directed graphs, not about whole directed graphs. Therefore, in this paper, we suggest a theorem 3.1 to get the transitive subgraph preserving topological information, introduce the explicit rule what edges are selected, and practical algorithm.

3.1 A transitive directed subgraph preserving corresponding topology

For a given directed graph, there are various ways to get a transitive directed graph by adding or eliminating edges. We suggest a way to get a transitive directed graph which preserves corresponding topology of a given directed graph.

Theorem 3.1. *For a directed graph G, let $G' = (\Phi \circ \Psi)(G)$. Then, G and G' have the same corresponding topology.*

Proof. By the lemma 2.18, G' is a transitive directed graph. Restricting the domain of the map Ψ to $\mathcal{G}_X^T$, Ψ is bijective and it is the inverse of Φ by the theorem 2.20. That is, we may denote

$$(\Phi \circ \Psi)(G') = G'. \tag{3.1}$$

Let us T and T' be the corresponding topologies of G and G', respectively. Then,

$$T' = \Psi(G') = \Psi \circ (\Phi \circ \Psi)(G) = (\Psi \circ \Phi) \circ \Psi(G) = \Psi(G) = T. \tag{3.2}$$

Therefore, G and G' have the same corresponding topology. □

3.2 Explicit rule to compute the transitive directed subgraph preserving corresponding topology

Now let us consider an efficient way to compute the transitive directed subgraph of a given directed graph. Computing each step of

$$(\Phi \circ \Psi)(G) \qquad G \in \mathcal{G}_X \tag{3.3}$$

is complex and requires high computational cost. Thus, it is not efficient. In this section, we suggest more simple and efficient way.

Theorem 3.2. *For a given directed graph $G = (X, V)$ on a finite set X, a directed graph G' is the subgraph of G which satisfies following rule: for distinct $a, b, c \in X$ with $ab, bc \in G$ but $ac \notin G$, then $ab \notin G'$. Then, G' is the same to $(\Phi \circ \Psi)(G)$.*

Proof. Let us take distinct $a, b, c \in X$ and assume $ab, bc \in G, ac \notin G$. Then,

$$\{a\} \subset a\downarrow, \quad \{a, b\} \subset b\downarrow, \quad \{b, c\} \subset c\downarrow, \quad a \notin c\downarrow. \tag{3.4}$$

Then, the smallest open neighborhood U_b of b is the smallest intersection of elements of basis which are containing b. i.e., let's observe $b\downarrow \cap c\downarrow$. Since $a \notin c\downarrow$, $\{a, b\}$ is not in U_b. Thus, a is not an adjacent point to b. Therefore, $(a, b) \notin G'$. □

Remark 3.3. By the theorem 3.2, we can compute G' without computing corresponding topology(cf., Figure 3.2).

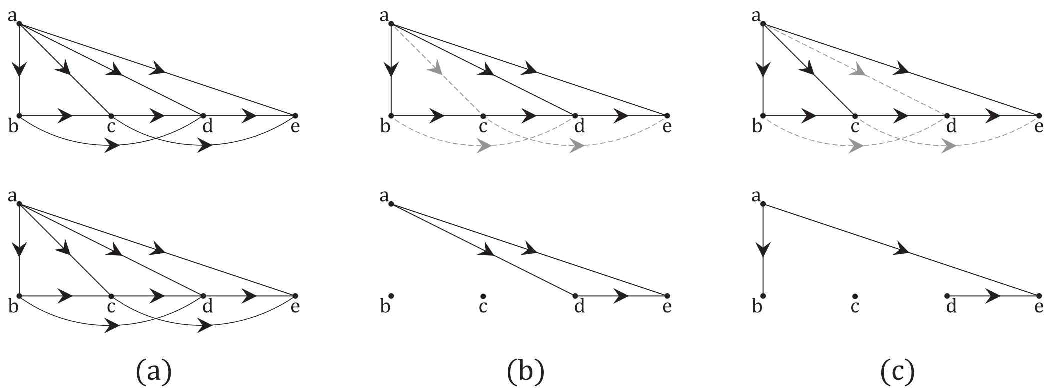


Figure 3: The graphs in the first row are the origin directed graphs. Dotted edges are not in the origin graphs. the graphs in the second row are the images of map $(\Phi \circ \Psi)$.

3.3 Computation algorithm via adjacency matrix

In the section 2.2, we said the adjacency matrix is a useful representation of a directed graph. Applying this, we introduce an algorithm. Let us see examples in order to design an algorithm computing G' without computation of corresponding topology T. Let $X = \{a, b, c, d\}$ and G with following adjacency matrix:

$$A(G) = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.5}$$

Then, we may know both edges (b, c) and (c, d) are in G, but (b, d) isn't. By the proposition 3.1, the edge (b, c) is not in G'. Hence, We may get

$$A(G') = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.6}$$

Now we suggest an algorithm without computing corresponding topology of given directed graph G.

Algorithm 2: Computation of $A(G')$

Result: The adjacency matrix $A(G')$ of G'

For given $A(G) \in \mathbb{R}^{N \times N}$

initialize $A(G') = A(G)$

for $1 \leq i, j, k \leq N$ **do**

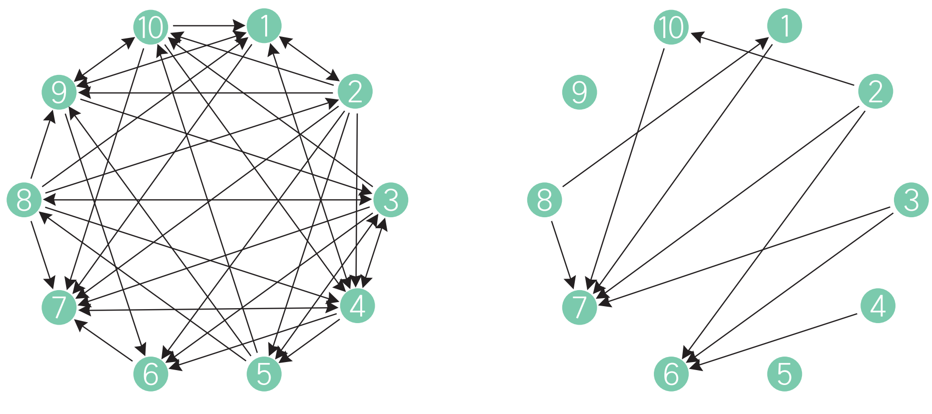
  if $[A(G)]_{i,j} = 1$ & $[A(G)]_{j,k} = 1$ & $[A(G)]_{i,k} = 0$ **then**

    $[A(G')]_{i,j} = 0$

  end

end

Example 2. Following figure is an output of algorithm written by python code. Seeing the figure, we may know pretty complex directed graph has simple transitive directed subgraph. Thus, for the directed graph with a lots of points, this algorithm will be useful when ones need to select points with some relation.



(a) (b)

Figure 4: (a) is a randomly generated directed graph with 10 points, (b) is the transitive directed subgraph by the algorithm 2

4 Transitive kNN subgraph of a kNN graph

We now apply the theorem 3.1 and 3.2 to kNN graph. First we set directions to edges of a kNN graph. Reminding the definition 2.23 in the section 2.5, in fact, we already announce the direction of edges. In addition, we may observe that kNN graph (X, G) has No loop and is asymmetric, in general.

observation 3. kNN graph is a directed graph. But, both acyclicity and transitivity are not guaranteed.

By applying theorems in 3, we get following conclusion of this paper.

Theorem 4.1. *For a given kNN graph (X, G), there is a topology defined by $\Phi(X, G) = (X, T)$. Furthermore, $(\Psi \circ \Phi)(X, G)$ is the transitive directed graph preserving (X, T).*

Application 1. In many areas, we can see datasets with partial order. Especially, so does the education area. Imagine a situation that study mathematics. For instance, almost people who learned 'sequence', 'series', and 'sequence of functions' have studied them along the strict order 'sequence, series, sequence of functions'. For personal service of education, we want to design a deep learning architecture which recommend personal curriculum and workbooks. In general, kNN graph has different degrees for each point, since out-degree is fixed by k but in-degrees are not. So we may interpret the points with higher in-degree as important points like hub. (i.e., the 7 and 6 points will be important points, in the figure 3.3). The following is an abstract of an recommendation algorithm applied by this paper. Practical experiment is impossible now because we can not stack a dataset for it. Thus, experiment is suppose to be a next work of ours.

Algorithm 3: A backward recommendation algorithm via topology

Input: Dataset with features

Output: A directed graph which is a recommendation path

1. Get a kNN graph via algorithm 1
2. Get the transitive directed graph via algorithm 2
3. Put in-degree as a score function of algorithm 1
4. Select points with local maximum degree in open set in algorithm 2
5. Take a subgraph connecting hub points
6. Optimize subgraph by selecting paths via deep learning technique.

5 Conclusion

In this paper, we propose and prove a relation of directed graphs and their transitive directed subgraph preserving topologies on a finite set. Furthemore, we suggest a simple algorithm computing transitive directed subgraph applying adjacency matrix. In the section 4, applying our suggestion to kNN directed graphs, we may get the transitive directed kNN subgraph with mathematical aspect. Thus, this paper has a novelty as a mathematical tool to analyze partial ordered datasets. However, there are several issues which we have to overcome to treating partial ordered datasets. For instance, how to get over Non acyclic issue of kNN. Therefore, we will study those issues during my Ph.D course.

References

[1] Jung Ho Park, Jeamin Park, Hyun-Young Choi, and Hyeon-Woo Jeung.: A Feature selection method based on topological aspect of feature manifolds. *Ksiam fall conference.* (2019).

[2] Evans, J., Harary, F., Lynn, M.: On the computer enumeration of finite topologies. *Comm. ACM.,* **10** (1967), 295-298.

[3] Carlos Mariuan: Finite topologies and directed graphs. *Proyecciones Journal of Math., Vol* **29** (2010)

[4] Benoumhani, M: The number of topologies on a finite set. *J. of Integer Sequence, Vol* **9** (2006).

[5] E. Sampathkumarachar: k-Transitive directed graphs and Topologies/ *International Mathematical Forum* (2010).

[6] A. Grigor'yan, Yong Lin, Y. Muranov, Shing-Tung Yau: Homologies of path complexes and directed graphs. *arXiv* (2012).

[7] X. He, D. Cai, P. Noiyogi: Laplacian score for feature selection. *Advances in neural information processing systems* (2006).

[8] V. de Silva, G. Carlsson: Topological estimation using witness complexes. *Eurographics Symposium on Point-Based Graphics* (2004).

[9] G. Singh, . Mémoli: Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition. *Eurographics Symposium on Point- Based Graphics* (2007).

국문초록

이 학위 논문에서는 유한집합 위에서의 위상들과 transitive 유향그래프의 일대일대응에 관련된 내용들을 상기한다. kNN 그래프는 데이터를 활용한 과학 분야에서 널리 사용된다. 우리는 kNN그래프의 in-degree와 out-degree를 기준으로 kNN그래프를 유향그래프로 해석할 수 있다. 하지만 일반적으로 kNN그래프는 transitive 유향그래프가 아니고 주어진 kNN그래프의 간선을 추가하거나 빼는 등 다양한 방법으로 transitive 유향그래프를 만들 수 있다. 이 학위 논문을 통해서 우리는 위상을 보존하는 transitive 유향그래프를 얻는 방법이 유일함을 증명하고 구체적인 방법을 제시한다. 더 나아가서 위상의 계산 없이 transitive kNN그래프를 구하는 계산 알고리즘을 제안한다.

주요어 : kNN그래프, transitive 유향그래프, 유한위상
학번 : 2017–28387