Traveling in Hakkaido

ZHU Weizhi
Department of Mathematics
HKUST

FAN Min
Department of Mathematics
HKUST

1 Introduction

Hakkaido, located at the north end of Japan, is the second largest island of Japan, which is well known for its successive beauty of the year. With its unspoiled nature, Hakkaido attracts millions of people every year. The major cities in Hakkaido includes **Sapporo** - largest city on the island, **Hakodate** and **Asahikawa** - two core cities in the south and in the central region respectively, **Muroran**, **Noboribetu**, **Abashiri** and many more appealing tourist destinations.

In this project, we regard the major 17 cities as nodes, and the route between cities as links. We try to define some kinds of energy on nodes, which could represent the degree of importance of the corresponding cities. By collecting data on tripadvisor.com, we build our Hakoaido tourism network. The following part is organized as follows: In section 2, we briefly introduce the data set and basic setup of tourism network; In section 3, we using hierarchical decomposition (Weinan E, Jianfeng Lu, Yuan Yao, 2013) to analyze the landscape of tourism network; In section 4, we using transition path theory (Weinan E, Eric Vanden-Eijnden, 2010) to analyze the 'classic' tourist route and introduce some of the modified versions to our problems.

2 Tourism Network and Data Set

In our Hakkaido Tourism Network, we select 17 major cities as nodes and there exists a link if and only if the shortest path of public transportation doesn't pass other nodes. We collect our data manually on official Japanese public transportation website [4]. The Figure below shows the graph, where we try to keep the relative position for better visualization.

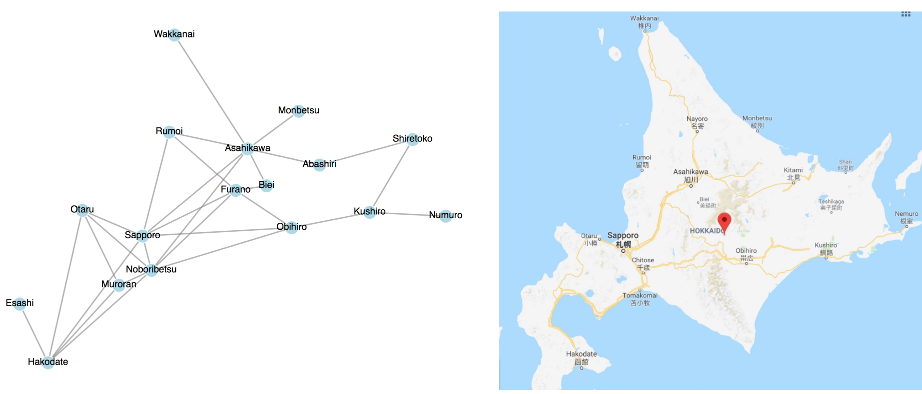

Figure 2.1: Left : Graph of Hakkaido Tourism Network, Right : Hakkaido Map from Google

Besides the graph, we also need to define a score function on nodes. We collect the score data manually from [3], through this project, we will use the average number of reviewers for each cities as score.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

3 Landscape of Tourism Network

3.1 Hierarchical Decomposition of Nodes

Generally speaking, hierarchical decomposition technique (Weinan E, Jianfeng Lu, Yuan Yao, 2013) is used for exploring the structure of network in order to uncover relationship between each potential communities. The methodology behind is to define energy function on nodes and examining the critical points by gradient flow between connected nodes. The **index k-critical nodes** here stands for the nodes of 'hub' but **in different level**.

The most important theory for hierarchical decomposition is given as follows,

Theorem 3.1 (Node Decomposition)**.** *Suppose graph $G = (V, E)$ and energy function on nodes f are defined, then V admits the following decomposition,*

$$V = \mathscr{B}_0 \bigcup \bigcup_{x \in \mathscr{C}_0} \mathscr{A}_0(x) \tag{3.1}$$

Where for $k \geq 1$,

$$\mathscr{B}_{k-1} = \mathscr{B}_k \bigcup \bigcup_{x \in \bar{\mathscr{C}}_k} \mathscr{A}_k(x) \tag{3.2}$$

Here $\mathscr{A}_k$ is the attraction basin of local minima restricted on the k-1 the boundary set $\mathscr{B}_{k-1}$ and $\bar{\mathscr{C}}_k$ is the set of nondegenerate index-k critical nodes.

Details of such concept could be found in the paper. Roughly speaking, decomposition 3.1 gives us the hub of the nodes, $\mathscr{C}_0$ and the most 'loyalty' nodes to the hub x, $\mathscr{A}_0(x)$. The rest of nodes $\mathscr{B}_0$ are the 'bridges' between different hub. Furthermore, decomposition 3.2 allows us to see the similar structure of $\mathscr{B}_0$. For example, in a social network, two gangs may be connected by an intermediary agent, others who want to make a deal with gangs must be introduced by the agent, which is the local minimum of $\mathscr{B}_0$.

We implement the Fast search of nondegenerate critical nodes algorithm in Python, which could be found in my github page(Soon).

3.2 Hakkaido Tourist Network Decomposition

Now we apply the Fast search algorithm to our Hakkaido touriest network. Here we let energy function $d = -log(scores)$.

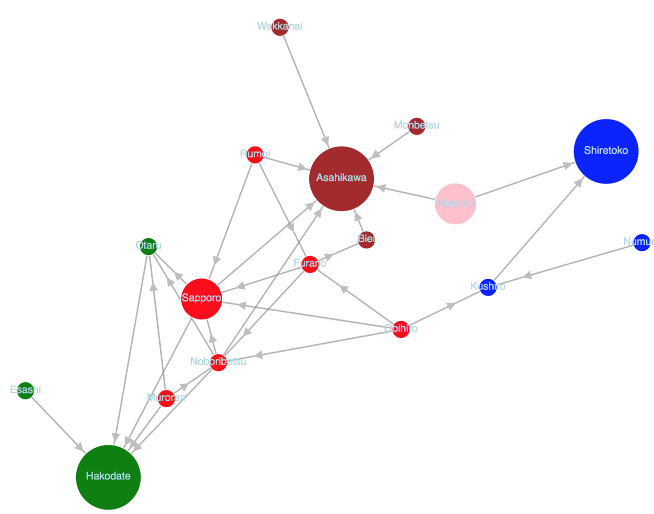

Figure 3.1: The radius of point corresponding to index. The biggest three nodes are the local minimums, while the middle nodes are index-1 critical nodes and others are attraction basins for each critical nodes. Nodes have same color means they share the same attraction basin.

As we seen in the figure3.2,

1. the local minimum of the graph are three cities, Hakodate, Asahikawa and Shiretoko. As be pointed in some introduction of Hakkaido, for example in Wiki, Hakodate and Asahikawa are the two core cities in south and centric area. Therefore, we are not wonder why it becomes a local hub in the tourism network. Besides, Shirekoto is regarded as a local hub in the network due to it's in a remote area, only two small cities connect it.

2. the index-1 critical points are two cities, Sapporo, Abashiri. Sapporo is capital of Hakkaido while in our model, it connects two core cities - Asahikawa and Hakodate. It's obvious that (Hakodate, Sapporo, Asahikawa) and (Shiretoko, Abashiri, Asahikawa) are minimal energy pass.

3. the index-1 critical points are the bridges between two communities. In tourism network, if one wants to travel from Hakodate to Asahikawa, people will prefer the minimal energy path, then it's reasonable for merchants to increase the price of local accommodation and for government to provide more public transportation.

4 Transition Path in Tourism Network and its Modification

4.1 Transition Path Theory

Generally speaking, tools of transition path(Weinan E, Eric Vanden-Eijnden, 2010) are aimed at understanding the mechanism and rate of specific reactions in the system, that is, transitions between any particular states or group of states of interest. To be more specific, we are interested in the reaction trajectories flow from the reactor(begin) to the product(end). There are some important values used for the explanation of reaction process, such as

1. **Transition Probability Matrix**, $P = D^{-1}W$. Here we define W to be the adjacency score matrix, which simply means the transition probability from a node to another (linearly) depends on the score of possible destination.

2. **Stationary Distribution**, $\pi P = \pi$.

3. **Committor Function**, $q(x) = P_x(\tau_A < \tau_B)$, that is the probability a trajectory visiting node x be 'reactive', i.e. that this visit occurs while the process is on its way from A to B. The (discrete) committor function could be solved by a (discrete) Dirichlet boundary problem.

4. **Transition Flux on Edge**, $J(x, y) = \pi(x)(1 - q(x))P_{xy}q(y)$, that is the joint equilibrium probability (after the process become stationary) that the trajectory visits states x and y consecutively and that it be reactive.

5. **Effective Transition Flux on Edge** $J^+(x, y) = max(J(x, y) - J(y, x), 0)$, that is the probability current of the reactive trajectories that indicates how, on average, reactive trajectories flow from A to B. Note that only one of $J^+(x, y) and J^+(y, x)$ could be large than 0, while it's also possible both are 0 when there is no such trajectory flow satisfy conditions.

6. **Transition Flux on Node**, $T(x) = \sum_{y \in V} J^+(x, y) = \sum_{y \in V} J^+(y, x)$, that is the in-nodes effective flux (or out-nodes effective flux).

Next, we will use these index to analyze the tourism network.

4.2 Hakkaido Tourism Network Transition Flux

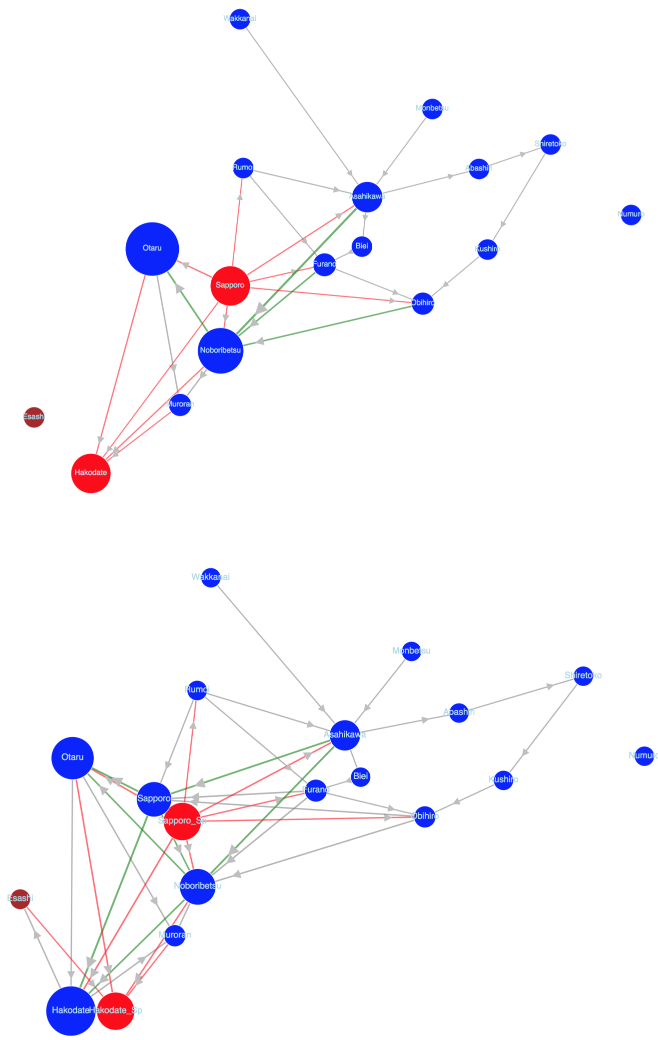

Figure 4.1: Sapporo to Hakodate : The red nodes indicate the origin and the end of a journey. The radius of nodes represents the effective flux through it and the thickness of line represents effective flux on corresponding edge where the arrow indicates the direction. Note that the red nodes and links are re-scale due to its extremely large value, which has no useful meaning for us. The green links represents high flux edge. Top : Non-split tourism network flux expression ; Bottom : Split tourism network flux expression.

When we directly apply transition path theory to the tourism network, with starting at Sapporo and ending at Hakodate. We want to capture the main travel route and cities in classic **'South Hakkaido Traveling'** . The flux expression capture some character, for example Otaru, which is famous for its Otaru Canal, has maximum flux-through nodes in Sapporo-Hakodate Journey. (We will give a more specific analysis towards Sapporo-Hakodate flux figure later.) However there are some **crucial problems** occur:

1. Comparing with other links, the link connect the origin and end have extremely large value. (In the picture we re-scale the red subjects for visualization consideration.) Same situation happens for the origin and end node. But actually, the high flux fails to reveal the importance of the city from tourism view, but only means it's starting point or ending point.

2. Transition path don't allow a visitor passing by the origin again. However, as we've seen in Section 2, Sapporo is indeed a hinge between two core cities.

3. Similarly, Transition path don't allow a visitor passing by the end before he end his journey. In the top of figure 4.2, the brown node, Esashi, has zero-flux since its only neighbor city is Hakodate.

In order to fix these problems, we introduce the split-strategy.

4.3 Split Nodes Strategy

Our idea is simply split the origin and end nodes into two parts. For example, we split Sapporo into two parts : Sapporo-Sp and Sapporo. The Sapporo-Sp is only the symbol of origin of the journey, while the Sapporo is a real city for traveling. Intuitively, we could explain Sapporo-Sp to be the airport, one could go to Sapporo for traveling or directly go to other cities without affecting our analysis towards Sapporo as a tourist attraction, since we do not regard the flux of Sapporo-Sp as flux of Sapporo. Similarly, we split Hakodate into two parts, visitors going to Esashi could be captured from our model now.

One detail we want to mention shortly is we have to set the transition probability between split points and others. In our model, we just give the split point a score and derive the transition probability. However, the score could be tuned by cross-validation based on some known information, for example stationary distribution.

As we could see from the bottom figure 4.2, we can capture many interesting information for Sapporo-Hakodate Journey, we list three of them for example:

1. The Sapporo-Hakodate Journey is basically traveling around Otaru(canel), Noboribetsu(spring), Asahikawa(core city, penguin zoo), Sapporo(capital) and Hakodate(night scene). Although Asahikawa have larger score than Noboribetsu and Otaru, but in the Southern Hakkaido Journey, people will prefer more about the later two.

2. According to the direction of main effective transition flux on Asahikawa-Sapporo, people prefer to travel like Sapporo - (Rumoi, Furano, Obihiro...) - Asahikawa - Sapporo. Similarly analysis to Otaru, if one go to Otaru, the effective flux tell us more people will go southern, which means they are less likely to go Asahikawa again.

3. Seldom people will travel to the north cities if their destination is Hakodate.

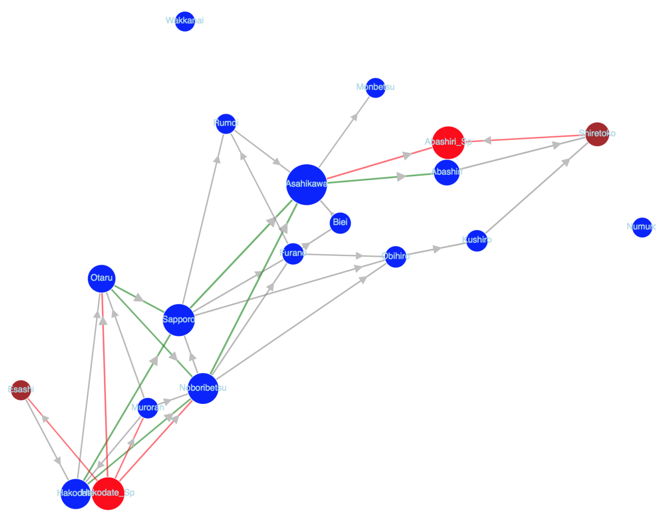

Figure 4.2: Hakodate to Abashiri : The red nodes indicate the origin and the end of a journey. The radius of nodes represents the effective flux through it and the thickness of line represents effective flux on corresponding edge where the arrow indicates the direction. Note that the red nodes and links are re-scale due to its extremely large value, which has no useful meaning for us. The green links represents high flux edge.

We also do lots of experiments on other choice of classic journey, for example Hakodate - Abashiri Journey, which travel from south to north. The result is shown in Figure 4.3.

4.4 Other Possible Modification

In fact, we could also further split each node into two parts : when visitors arrive in a city, the visitors will make a decision, to stay for traveling or to directly leave to another city. The idea behind these is we want to judge whether a city with high flux is due to its popularity or just as a traffic hinge. Besides, we could block some cities together and use this idea to consider the transition path when the origin and ending are same : we could separate the whole journey into two process, away process and back process. In the away process the destination is group of cities having distance larger than a value, while the back process is the opposite. We leave these for future exploration.

Acknowledge

We are grateful for the guidance of d3js by Xavier.

References

[1] Weinan, E., Jianfeng, Lu, and Yuan, Yao. "The landscape of complex networks." arXiv preprint arXiv:1204.6376 (2012).

[2] Vanden-Eijnden, Eric. "Transition path theory." Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1 (2006): 453-493.

[3] https://www.tripadvisor.com

[4] http://www2.jrhokkaido.co.jp/global/chinese/ttable/index.html