Uncovering Political Promotion in China: A Network Analysis of Patronage Relationship in Autocracy

Zhengxuan Wu*, Jason Luo*, Sherine Zhang*

*authors contributed equally to this work

Abstract—**Understanding patronage networks in Chinese Bureaucracy helps us quantify promotion mechanism underlying autocratic political systems. Although there are qualitative studies analyzing political promotions, few use quantitative methods to model promotions and make inferences on the fitted mathematical model. Using publicly available datasets, we implement network analysis techniques to advance scholarly understanding of patronage networks in autocratic regimes, using the Chinese bureaucracy as an example. Using graph-based and non-graph-based features, we design three studies to examine drivers of political promotions. We find that careers of politicians are closely associated with their genders, home origins, and positions in the patronage networks.**

I. INTRODUCTION

Interacting with others and forming connections are important skills among top executives in large companies and government institutions. Previous literature qualitatively demonstrated that informal connections help employees go around with formal constraints in large institutions [1]. How the patronage network will affect promotions of politicians? What features of the network play important roles in promotions? These are the questions on which we can make several different arguments.

Previous literature mainly focuses on the qualitative measurements of effects of patronage among China's political elites. They often came from limited insider sources [2]. As a result, these studies often end up with theoretical speculations and thus lack statistical inferences. Although researchers often use the term "network" in political research [2], only a few scholars have applied actual social network analysis (SNA) [3].

We combine network analysis with statistical techniques, specifically with the goal of studying promotions in government institutions. The limited studies in patronage networks mainly focused on local network features and tried to use network analysis to study current state of Politburo behaviors. They did not try to use statistical tools to infer future political characteristics based on those features [4]. Our study draw lines across network analysis, statistical analysis and political promotions in autocratic regimes. The insights will deepen our understanding of the structures of autocratic institutions, the patronage mechanism and the promotion process.

In this paper, we apply social network analysis to the Chinese Political Elite Database (CPED) dataset [5]. Our research begins with the constructions of three categories of networks - home origin, overlap-based patronage and promotion-based patronage. We applied linear regression to find correlations between the promotion history of politicians and their direct superior.

Then, we use regression models to find correlations between promotions and both external and network features among Chinese politicians in our dataset. Our exploratory variables include basic features from biographies and advanced features from networks, such as node level features and structural level features.

Finally, based on the fitted results, we interpret the fitted parameters from those models and try to discover local features, external features and network features that are highly correlated with an politician's promotion in career. We then apply machine learning techniques to make predictions using these features.

II. RELATED WORKS

A. Patronage Effect

Patronage networks play an major role in regimes around the world, especially in China [6], [7], [5], [1], [3], [4]. Previous studies have shown Guanxi, which means "patronage network" in Chinese, plays

a central role in promoting employees in large institutions, such as private or public-owned companies and government [5]. Likewise, researchers state that the patronage network of the formal Prime Minister of China, Jiabao Wen was worth billions [8].

Similarly, studies have shown that the career promotions of politicians are closely related to their relations with colleagues in the patronage network [5]. For instance, researchers in India pointed out that having close relationships with one's superiors would increase one's chance of promotions [9]. Other studies found that forming personal connections within institutions would help employees get promotions and work around constraints [5]. We believe that the effect of the patronage network on promotions could be longitudinal. In other words, the promotion trajectory of a politician would be positively associated with the promotion trajectory of his or her closely related connections in the patronage network.

B. Social Network Analysis

Network Analysis has been used in different fields of research, including information science [10], kinship analysis [11], online service recommendations [12], co-author citation analysis [13], gene-disease analysis [14] and online social network analysis [15]. For example, researchers identified studies with similar topics by applying graphical analysis on the co-author citation network. Likewise, by applying network analysis to gene-disease network, they learned how genetic and environmental factors, such as drugs, contributed to diseases [14].

In social science, researcher used network analysis to study community structures and predict structural appearances by incorporating machine learning techniques [16], [17]. For instance, researchers proved that certain community structures existed in the selected networks by studying the interactions between their components [16]. Similarly, recent literature stated that edges between nodes in the network could be predicted by combining machine learning and network analysis in social network [17]. Current studies also showed that graphs could be auto-generated by applying generative models with networks analysis [18].

C. Inference With External Features

The promotion of a politician is determined by many external factors, such as one's work experience, genders, leadership skills, and economic growth of the city [19]. For example, studies proved that there existed a gender bias in political promotions [20]. Particularly in work-space promotions, researcher found that, although the wage differences were not huge, female candidates had significantly lower probabilities in promotions compared to the male candidates [21].

Home origin is another external feature. Interpersonal distance is the psychological term for measuring how close people are [22]. Studies proved that people who came from the same country would have stronger bounds in social networks [23]. In this case, we define two people to have the same home origin if they come from the same city.

D. Inference With Social Network Features

The patronage network will affect one's promotion [3]. Previous studies have shown that faction, for example, will affect a politician's career. There are also correlations between faction network, schooling network and home origin network [3]. We will extract different network features from all the networks we built, and test their correlations with politician's trajectories. It will have node level features, including degress, n-hop features, and structural level features, including role and motif detections.

III. DATASET

We will be using the Chinese Political Elite Database (CPED), a large biographical database that contains extensive demographic and career information of over 4,000 key city, provincial and national leaders in China since late 1990s [5].

For each leader, the database provides information about the time, place, organization, and rank of every job assignment listed in one's curriculum vitae, which is collected from government websites, yearbooks, and other trustworthy Internet sources. The author matches each city-year spell in the panel data set with a city secretary and a mayor. In cases where multiple leaders held the same post within a given spell, the person with the latest entry date is chosen.
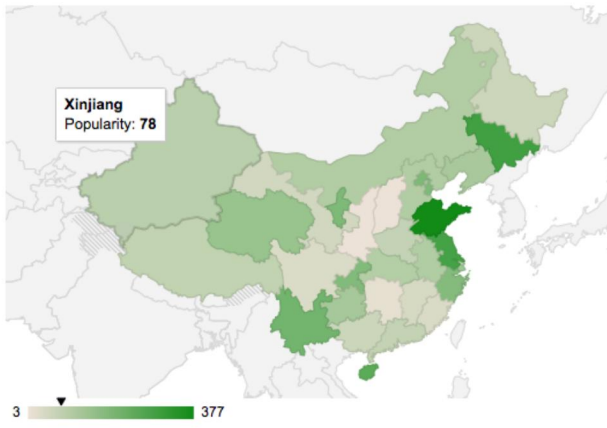
Fig. 1. Heatmap of hometown origin of all politicians in the dataset: Xinjiang has 78 politicians.

Figure 1 shows the distribution of politicians in the dataset according to their home origins. According to the results, we can see that the politicians are distributed across the entire country. On the other hand, most of them are from the east coast, whereas handful of politicians come from the middle of China. This maps with the population distribution in China. Table 1 shows details about the dataset, including the counts of politician, province and city. It contains the total count of data points in our work experience table.

Item	Count
Politician	4057
Province	32
City	389
Work Experience Data Points	62742

TABLE I
DETIALS ABOUT THE POLITICIAN DATASET

IV. NETWORK DEFINITION

This section describes how we use the plain text dataset of Chinese politicians to make our networks and graphs. We construct three main networks to simulate the patronage network in real life, including hometown origin network, work experience overlap network and promotion network. The dataset encodes 4057 politicians in total, which are considered in all three graphs as nodes.

A. Home Origin Patronage Networks

The hometown network is highly clustered based where they come from. For each city group, all nodes are completely connected. They all have the same degrees.
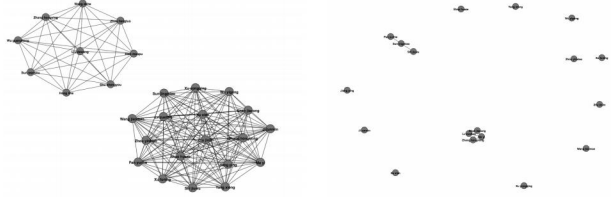

Fig. 2. Left graph is the networks of politicians based on their home origins. This network is for all the politicians from two cities, Hangzhou(bottom right) and Shangrao(top left). Right graph is the networks of politicians based on their home origins and work experience. All these politicians are from Hangzhou and share work experience.

1) Edges: We have 59306 edges within the network. Each edge represents that two politicians come from the same city. We have two different networks constructed. For the first graph, two nodes are connected if they are from the same city. For the second graph, two nodes are connected if they are from the same city and they have worked together in the same department at some point in their lives.

2) Visualization: To avoid unreadable visualization, we plot hometown networks for two cities as an example, shown in Figure 2. All the nodes within the clusters are all inter-connected. The network on the right is less dense as we restricted the network to only have an edge between those who have worked together. As a result, many nodes are not connected. Some groups form strong connections, as those politicians are from the same city and also share working experiences.

B. Overlap-based Patronage Network

We next construct a directed, weighted graph based on overlapping work/school experiences of the 4,057 political leaders.

1) Edges: We have 655,769 directed, weighted edges. Each edge in the network indicates the existence of at least one overlapping work or school experience between the two leaders. We consider an overlap if two have to work together for at least six months in the same municipality in the same province. if multiple overlaps are founded, we encoded into to edge weights, given by the total time of two working together. The direction of an edge is determined by comparing which cadre is senior to the other in terms of their average cadre level during the time periods they work together.

3

2) Visualization: The node distribution plot in Figure 3 suggests that most nodes have degrees between 100 and 1000. The right graph in the Figure
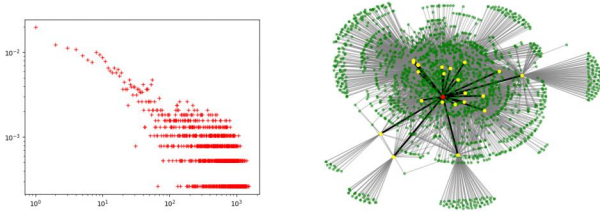

Fig. 3. Left graph is node Degree Distribution. Right graph is two-hop Neighbors Sub-graph of A Random Node 3507

3 shows the two-hop neighbors of a random node 3057, named Zhao Zhuping, currently the head of a district in Shanghai, whose rank is equivalent to a mayor in U.S. He has 22 patronage relationships under our definition, way lower than average in our data-set. This is partly because he is relatively young and serves only a moderate position in the bureaucratic system.

We plot node similarity graph between this node 3507 and all others to see how typical this node is and what roles other node have. The plots in Figure
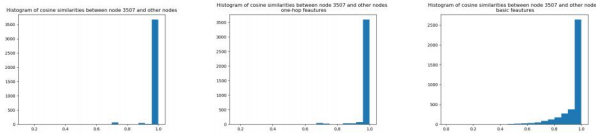

Fig. 4. Left graph is the node similarity distribution between Node 3507 and All Other Nodes (Two-hop Features. Middle graph is the node similarity distribution between Node 3507 and All Other Nodes (One-hop Features). Right graph is the node similarity distribution between Node 3507 and All Other Nodes (Basic Features)

4 show that most of the nodes are almost identical to node 3507, possibly because the network is so dense that the two-hop feature aggregation was able to capture the entire graph. We also plot one-hop and basic similarity distributions. The two plots show very clearly that the vast majority of nodes are very similar, which is suggesting the sub-graph we drew earlier for Node 3507 is highly representative across all nodes in spite of their differences in node degrees.

C. Promotion-based Patronage Network

Our third graph models patronage relationship network using political appointments among leaders.

1) Edges: There are in total 3905 directed edges in the network. An edge exists between two nodes when there exists a patron-client link. A patron-client link from A to B forms when A is promoted by B. In our case, two nodes are connected if one node is promoted by another. Specifically, we look at promotion from rank level 4 to 5 based on the CPED dataset. The promotion between level 4 and 5 is considered as a milestone in one's political career. [5]. The edge goes from the node being promoted to the its promoter. This network is not weighted because we are only considering one promotion.
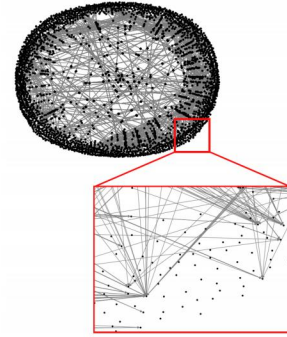

Fig. 5. Visualization of entire patronage network on the top, with a subset of it zoomed in to show clear edges between nodes.
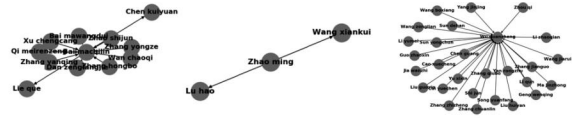

Fig. 6. Subgraphs with randomly chosen nodes and their edges.

2) Visualization: To construct Figure 6, we randomly selected 30 nodes and drew out their egonets. We included all of the node's one-hop neighbors. Based on the outputs, we observe three most common structures, as shown in Figure 6.

The left subgraph shows that the node has two outgoing edges and many incoming edges. In our graph, a node can have at most two out-neighbors, because the dataset is constructed in a way that a leader can have at most two direct promoters. The node in the middle subgraph has no incoming edges, possibly because it has not reached a level that can promote others. Another common structure is shown in the right subgraph, namely, the node has only incoming edges but it has a great number of them. As shown in Figure 7, the cosine similarity between node 1568 and other nodes reaches spikes between
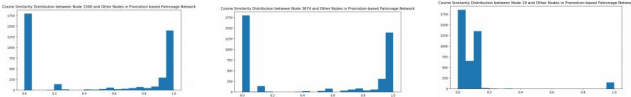
4

Fig. 7. Left graph is the node similarity distribution between Node 1568 and All Other Nodes (Two-hop features). Middle graph is the node similarity distribution between Node 3674 and All Other Nodes (Two-hop features). Right graph is the node similarity distribution between Node 19 and All Other Nodes (Two-hop features).

0.0 and 0.5, and 0.95 and 1.0, meaning that the most of the nodes are either identical to this node or completely different from this node. There are approximately 1400 identical nodes.

The cosine similarity between node 3674 and other nodes has a similar distribution. There are approximately another 1400 nodes identical to node 3674, apart from the 1400 nodes identical to node 1568, which adds to a total of almost 3000 nodes.

The cosine similarity between node 19 is different from the previous two. It has spikes between 0.0 and 0.05, and 0.1 and 0.15. Unlike our observation from Figure 6, there does not exist many identical or even similar nodes as node 19. Approximately 200 identical nodes are found.

V. METHOD

A. Study 1: Predicting Cadre Final Rank Using Network Features

Beyond descriptive inference, our first attempt for predictive analysis is to look at the extent to which network features predict final career result for leaders. Our baseline model is as simple as follows:

$$\text{Rank} \propto \alpha + \beta V_{NodeFeatures} + \gamma C_{covariates} + \epsilon \quad (1)$$

The outcome final rank is a leader's cadre level, which is one's political level in Chinese government, at 2015. For a node feature vector, we use node's feature vector up to two-hop aggregation, which is a vector of length 27. We control for one's birth year, year of joining the communist party, and year of promotion to municipality-level, which is rank 5, to adjust making comparisons across different stages of cadres' life and career.

B. Study 2: Detecting Political Factions

For this study, we analyzed the effects of gender and home origin on political factions. We used all three networks in this study. For the overlap-based

patronage network, we down-graded the network by removing the edge weights and directions. We assume that if two politicians have worked together, they are closely related.

1) Gender Effect: For each network, we define the *total degree* of each nodes as the sum of in-going edges and out-going edges.

$$D_i = \sum_{in} e(j, i) + \sum_{out} e(i, j) \quad (2)$$

All the edges are undirected. For each graph, we define the proportion of nodes given a total degree as the number of nodes with a given total degree divided by the total number of nodes in the network. We look at gender differences in terms of the total degrees of nodes in the overlap-based patronage network and promotion-based network.

We also define the rankings for all the politicians to be integers ranging from 1 to 9, with 9 being the highest level, the national leaders (Zhen guo ji). We define politicians with level above 7 as a *high rank politicians*, or *political elites*. For each node in our networks, we define a term *average ranks of neighbors* as the average ranking of all 1-hop neighbors of the node.

$$NbrRank_u = \frac{1}{|Nbr(u)|} \sum_{v \in Nbr(u)} Rank(v) \quad (3)$$

We look at the differences of the counts between male and female high rank politicians and compare the differences of average ranks of neighbors among male and female politicians.

2) Home Origin Effect: We use overlap-based patronage network to analyze home origin effects on political networks. After the network is down-graded, we embedded all nodes into vectors using $node2vec$ [24].

$node2vec$ works by carrying out a number of random walks from each node in the graph, where the walks are parameterized by p and q. In order to validate that people with same home origin have stronger connections in the graph, we used the BFS approach for $node2vec$ by setting the exploring parameters. More precisely, after having just traversed the edge from node node t to node v, the unnormalized transition probability of travelling

5

from node v to a neighboring node x is given by:

$$\alpha_{pq}(t,x) = \begin{cases} 1, & \text{if } d_{tx} = 0 \\ 1, & \text{if } d_{tx} = 1 \\ 2, & \text{if } d_{tx} = 2 \end{cases} \quad (4)$$

We sampled 10% of nodes from the original graph before running $node2vec$. We calculate two lists of scores for each node: *In Set Similarity Scores* and *Out Set Similarity Scores*. The scores are calculated by taking the dot product of the embedded vectors of two nodes. Two nodes are considered as an *In Set* pair if they are from the same province. For example, if two politicians are both from Shanghai, their similarity score, which is the dot product between two node vectors, would be added in to the *In Set Similarity Scores* for that node. Then, we explore at the province level. For each province, we iterate through every politician from that province and concatenate their *In Set Similarity Scores* and *Out Set Similarity Scores*. We then define the average scores within the province as the average of each list as following:

$$score_u = \frac{1}{|province_u|} \sum_{p \in province_u} score_p \quad (5)$$

$province_u$ is the set of politicians in that province. $score$ can be either within or inter-province scores. We compare those two scores to validate the effect of home origin in political networks.

3) Bridging Candidates Effect: After assigning nodes to different groups, we define the groups as cliques. For each node, we define within-clique edge count as the count of edges that a node has, to the nodes that are within the same clique as itself. Similarly, we define the term between-clique edge count as the count of edges that a node has, connecting to the nodes that were in different cliques from itself.

We investigate the correlations between the within-clique edge count, between-clique edge count and rankings of politicians.

C. Study 3: Hubs and Authorities

In this study, we use the Hyperlink-Induced Topic Search (HITS) algorithm to explore the hubs and authorities among two of our patronage networks, namely, overlap-based and promotion-based networks.

HITS algorithm is an algorithm used to analyze web links. It defines two types of Web pages and calls them *hubs* and *authorities*. The *authorities* web pages are usually prominent sources for a specific question or content. These pages are given high authority scores. On the other hand, the *hubs* are those that link to authority pages and act as a guide to other authority pages, usually those with a high authority scores. These Web pages are given high hub scores [25].

From the dataset, we extract the rank that every political leader is at the *end year* of this dataset. The *end year* is defined as 2015 or the year they retire. We defined the rank as *the final rank* of the politician. Every political leader has a final rank ranging from 0 to 9. We then calculate the hub and authority scores for both networks, and plot the scores against final ranks, trying to identify correlation. Scatter plot is used to show a general structure. For every rank, we take the average score of all political leaders of that rank and plot the average score on top of the scatter graph. Then we fit a linear line of the average scores against final ranks.

VI. RESULTS

In this section, we will discuss results we found for our studies.

A. Study 1: Predicting Cadre Final Rank Using Network Features

Our preliminary OLS regression model yields a R^2 of 0.625, suggesting 62.5% of the variation in cadre final ranks are captured by our explanatory variables. Then we evaluate our baseline model's prediction power on in-sample and out-of-sample performances. For out-of-sample, we hold out 10% of the data, estimate model on the other 90%, and test model on the 10% for cross validation. Thus, our out-of-sample prediction is averaged over all 10 hold-out sessions.

	Accuracy
In Sample Predictions	0.721
Out of Sample Predictions	0.718

The closeness of the accuracy results suggests that we are not over-fitting and we have a pretty good baseline results to start. We now shift to some

more complex model. First, we try Ordinal Logistic Regression. The model specification is as below:

$$P(Y_i = j) = \frac{exp(\tau_j - X\beta)}{1 + exp(\tau_j - X\beta)} - \frac{exp(\tau_{j-1} - X\beta)}{1 + exp(\tau_{j-1} - X\beta)} \quad (6)$$

Assuming,

$$Y \sim \text{Multinomial}(1, \pi)$$
$$Y^* = X\beta + \epsilon$$
$$Y^* = Y(\tau_j)$$
$$\epsilon_j \sim_{iid} \text{logistic}$$

Here Y_i is the rank outcome for each leader, and j is from 0 to 9 (10 levels of cadre rank), and X is feature vector up to two-hop aggregation and years of birth, joining party, and promotion, same as those in OLS. The table below gives results:

	Accuracy
In Sample Predictions	0.666
Out of Sample Predictions	0.649

Results above suggest our multi-nomial (ordinal) logit model doesn't work better than pooled OLS.

The next step is to use early-stage patronage network information to predict leaders' career outcome in the end. Specifically, we experimented with using network structures up until 2005-07-01 to predict cadre ranks ten years later, on 2015-07-01.

Using similar specifications as above for OLS and Logit, except for holding off rank information post 2005, we obtained the following results.

	Accuracy
In Sample Predictions (OLS)	0.690
Out of Sample Predictions (OLS)	0.683
In Sample Predictions (Logit)	0.725
Out of Sample Predictions (Logit)	0.720

Above results show that we have about 2.5 percentage points increase in accuracy from the OLS model, and almost no gain for the logit model. This suggests node features covering 2005 to 2015 add no gain to our prediction power (at least via the node2vec method we represent them). Network structures after 2005 are even noisier in predicting career outcome in 2015.

To make sure our test does not suffer from model specification or baseline results, we further regress outcome solely on node feature variables, without adding any other co-variants and outcome on nothing with the interception only. The former model yields an in-sample accuracy of 0.581 and out-sample of 0.575, and the empty model has an in-sample of 0.386.

B. Study 2: Detecting Political Factions

1) Home origin distribution of top rank politicians: According to the hometown network, we found that some provinces have larger population of politicians, some less. Combining with the ranking information, we plot out the home origin distribution, in province level, for top ranked politicians.

We found that all the top ranked politicians are from northeastern regions of China, and they are from regions close to the pacific east. These results are similar to the gross domestic product distributions across the country [26], shown in Figure 10 ($t = -2.11, p = 0.03$).
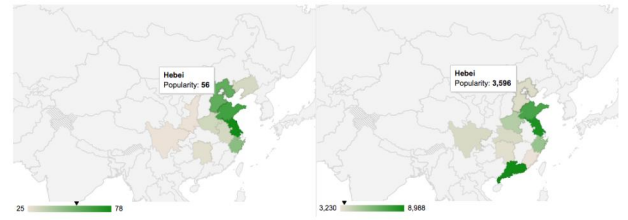

Fig. 8. Left is the top 10 home origin (province) distribution of the top politicians: for example, we have 56 politicians ranked in minister level from Hubei Province; Right is the gross domestic product distribution of 10 top rank provinces: for example, Hubei has 3596 dollars as its gross domestic product.

2) Gender And Connectivity: We calculated degree distributions for male and female candidates ($t = -2.12, p = 0.03$). Figure 9 showed that the average degrees of male and female politicians were close ($\mu_{male} = 341.81, \mu_{female} = 390.68$). The female politicians had higher average degrees. However, we have more high connected male politicians than female politicians.
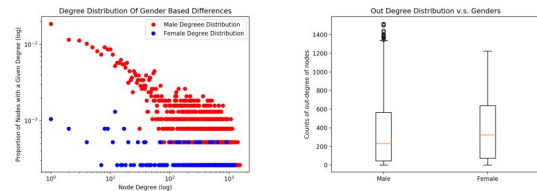

Fig. 9. Left is the degree distributions for male and female politicians. Right is the box plot for the degree distributions.

3) Gender And Rankings: We calculated the ranking differences between male and female politicians. Ranking distributions of male and female candidates are different. Similar to what we had for the degree distributions, the average rankings of female politicians was higher than that of male. However,
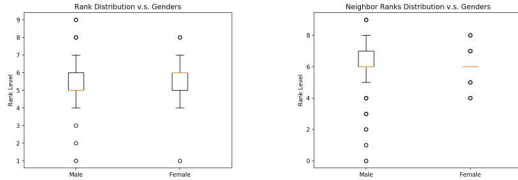
Fig. 10. Left is the box plot for the ranking distributions for male and female politicians. Right is the box plot for the ranking distributions of neighbors for male and female politicians.

we have more high ranked male politicians than female politicians.

Likewise, we calculated the average rankings for the closest neighbors of male and female candidates ($t = 3.40, p = 0.0067$). The trend was the same.

4) Home Origin And Similarity: The similarity score from $node2vec$ showed politicians from the same home origin had tighter connections in general ($t = -1.19, p = 0.24$).
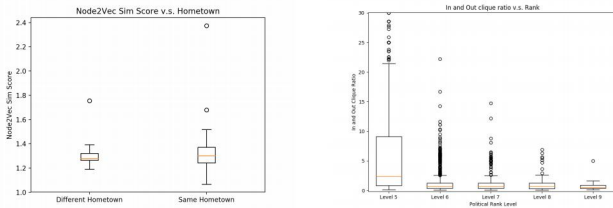

Fig. 11. Left is the similarity score distributions for different hometown and same hometown groups. Right is the With-in clique and Inter-clique clique ratio with respect to the level rankings of the politicians. Level 9 is corresponding to the level of the president.

In our case, hometown represented the province the politician came from. More than 70% of the cases, politicians from the same hometown had higher connectives in the $node2vec$ embedded space. We have politicians from $Shanghai$ stood out as an outlier. The average within hometown community similarity score was 2.37. This means politicians from $Shanghai$ tended to have stronger connections during their careers.

5) Within Clique And Inter-clique Connectivity: For this study, we defined that two politicians came from the same clique if they had the same home origin. Based on our (work experience) overlapping graph, we defined with-in clique connectivity as the count of edges going out from one node to other nodes who had the same home origin attribute. Similarly, we defined inter clique connectivity as the count of edges going out from one node to other nodes that had different home origin attribute. We

then calculated the with-in clique and inter clique edge count ratio for each node. We plotted the ratio against the ranking level of a node. We only took nodes that had ranking level greater than 5, since they were considered as high ranking politicians. We limited the ratio from 0 to 30, since for any nodes with ratio above 30, we could consider that they only had with-in clique edges.

In the plot, we have ranking levels ranging from level 5 to level 10. Level 5 is the equivalent level for the state Governor. Level 10 is the presidential level. We can see that there is a big jump in the clique ratio from level 5 to level 6 ($t = 10.69, p \leqslant 0.01$). We can see the from level 7 to level 8, there is no jump ($t = 0.71, p = 0.48$).

C. Study 3: Hubs and Authorities

We generate four graphs based on the two networks. Shown in Figure 12, the scatter plot for the Promotion-based Network indicates that different rank groups may correspond to hubs and authorities. However, the average score of every rank does not show such relationship. Both of the hub graph and authority graph show rather flat lines for the average score, which indicates that no significant correlation is found in the Promotion-based Network.

Fig. 12. Left is the hub scores as a function of final ranks of the political leaders based on the Overlap-based Patronage Network. Right is the authority scores as a function of final ranks of the political leaders based on the Overlap-based Patronage Network.

The Overlap-based Network, however, shows different structures. A surprising finding from the hub and authority graph is that they seem to be very similar. We explore the actual number of the scores and find that the hub and authority scores for almost every political leader are exactly the same before five decimals after the decimal point and only differ after that.
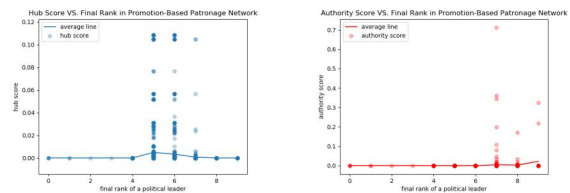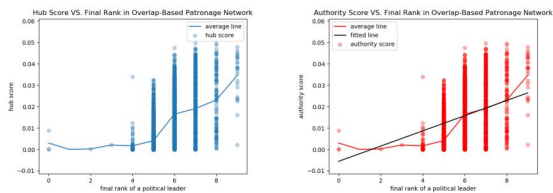
8

Fig. 13. Left is the hub scores as a function of final ranks of the political leaders based on the Promotion-based Patronage Network. Right is the authority scores as a function of final ranks of the political leaders based on the Promotion-based Patronage Network.

TABLE II
THE CORRELATION BETWEEN THE AUTHORITY SCORES AND THE FINAL RANK OF POLITICIANS IN THE NETWORK.

	Dependent variable:
	Authority Score
Final Rank Of A Political Leader	0.00355**
	(0.00064)
(Intercept)	-0.00556·
	(0.00342)
Observations	10
Note:	·p<0.15; *p<0.05; **p<0.005

Based on Figure 13, both of the scatter plot and the average line show some correlation between the final rank and the hub/authority scores. Because of the similarity of both graphs, we only fit one of the graphs and believe it will be almost identical for the other one. The fitted line is shown in black on the right subgraph of Figure 13, and the statistics is shown below in the table.

As shown in the table, there is a clear and significant correlation between the authority scores and the final rank of a political leader, indicated by the small p-value. We can say that the authority score increases as the final rank goes up for a political leader. Other statistics are reported as the following: $(F_{(1,8)} = 30.74$, RSE = 0.005821, Adjusted R^2 = 0.7677)

VII. DISCUSSION AND IMPLICATION

In our first study, we found that career outcome of political leaders in China is largely determined by patronage networks formed early in career, mostly even before promotion to municipality-level positions. It suggested that our network features up until 2005 adds strong prediction power to the model, implying that Chinese cadre's patronage relationships early in career may matter much more than

we previously thought. The strong and substantial implication here is not causal and we certainly do not claim exhaustive in capturing factors determining career outcomes with around 70% of accuracy in prediction. Nonetheless, our finding points to an interesting pattern that one's political career is largely arranged and destined by factors in her early years, be it overlapping work experience or other pre-determined features. Future research can do two things: (1) to figure out better ways of represent network features to make better predictions, e.g. Random Walks, Node Centrality, etc. Our current representation uses simple aggregation of neighborhood node basic features; (2) to explore mechanisms and factors that causally determine one's early career trajectory.

Our results from the home origin studies showed that the final ranking of a politician is highly correlated with the Gross Domestic Product (GDP) of one's home origin. This is not surprising, as we believe that higher GDP will bring a politician more resources and connections compared to a politician from a rural city in China.

We found that the average ranking of female candidates is similar to the average rankings of male. However, there were drastically more male politicians in high-ups. This results coincided the stereotype of the gender differences in political institution. Currently, we do not have a woman among all the highest level politicians in China. These results suggested that female politician would have lower chance to get promoted in autocratic countries, such as China. We found the same gender differences in degree distribution as well.

We studied the closeness in overlap network of nodes that are from the same home origin. We found more than half of the cases, politicians with the same home origin had higher connectives in the work overlapping network. People that were from the same hometown tended to work together. We found that *Shanghai* had a much higher score in this case, comparing to other cities. This suggested politicians from *Shanghai* focused more on patronage relationships in political promotions.

We found politicians with higher rankings tended to have more out-clique connections. There is a big jump between level 5 and level 6. Level 5 politicians work at a state level, whereas level 6 is above state level. Our model suggested that a politician forms more out-clique connections, especially after Level

6, which makes sense in that politicians work with people from other cities and provinces after they are promoted to a national level.

In study 3, we analyzed the hub and authority scores for both the overlap-based and promotion-based networks. Our plot shows no clear correlation between the scores and final ranks on the promotion-based networks. This is due to how the promotion-based network is set up. When constructing the network, we only considered a promotee's direct superior to be his or her promoter. Higher ranked politicians are not included, which explains the lack of correlation on this graph. However, a positive linear relationship is shown on the overlap-based network. As a politician's final rank goes up, the politician tends to be more central in the network, and thus have higher HITS scores, both in terms of having more important connections and in terms of having higher authorities.

Lastly but certainly not the least, our overall analysis of patronage network information from various aspects yields a great deal of understanding for career outcome of Chinese political leaders. Again, by simply looking at the early 10-year overlapping work experience, we were able to predict career outcomes ten years later, at an accuracy level of around 70%. This is simply amazing because the Chinese official line of cadre promotion is based on economic and governance performance, so does a major stream of literature argue. We show that patronage network, if not overwhelmingly, largely determines career ladder, and thus performance is either a second-order factor or a channel variable through which patronage factor translate itself into career outcome, i.e. politicians with better patronage relationships are assigned to regions and positions that are more likely to yield better economic performance.

VIII. Limitations

In this study, we did not exhaust the list of factors that could play a role in political promotions. In the home origin study, we only look at candidates that are from same city. The scope of a city might be too large for defining close relationships. Future studies may consider narrowing down the definition of home origin between politicians with a larger dataset.

Likewise, for the promotion network, we only consider the promotion between level 4 and level 5. Future researchers could explore the dataset to include patronage relationships incurred from every promotion during a leader's political career.

Further research can explore other methods in representing network structures into vectors, and conduct finer-grained research into mechanisms through which these patronage networks translate into career outcomes.

IX. Conclusion

In this paper, we have used network analysis techniques and statistical tools to understand the mechanism behind promotions in autocratic political systems. Specifically, we focus on the interpretations of the patronage network.

By using social network analysis on the three constructed networks, we found associations between the final rank of a politician and one's early stage patronage connections. We find that early stage development in the patronage network plays a more important role in one's political career compared to a more current stage. We have found the correlations between the gender and the final rankings, as well as the connectives. Our result suggests that politicians from same hometown tend to be much closer in the patronage network. On the other hand, we have found that in order to jump from state level position to above, one needs more out of hometown connections. We also find that hub and authority scores are direct mappings of political rankings in the patronage network.

In conclusion, our study presented a new approach to analysis patronage relationship in autocratic political system using publicly available promotion records. Combining network analysis and statistical analysis, we were able to quantify the promotion mechanism which was long thought to be mysterious in autocratic government. We found that our results coincide with theoretical speculations from previous qualitative studies, which came from limited insider sources. We provide a new way to further quantify the promotions of autocratic political systems using publicly available data.

Github Link
https://github.com/sherinezzzzzzz/224Wproject

REFERENCES

1 Xin, K. K. and Pearce, J. L., "Guanxi: Connections as substitutes for formal institutional support," *Academy of management journal*, vol. 39, no. 6, pp. 1641–1658, 1996.

2 Tsou, T., "1. prolegomenon to the study of informal groups in ccp politics," *The China Quarterly*, vol. 65, pp. 98–114, 1976.

3 Keller, F. B., "Moving beyond factions: Using social network analysis to uncover patronage networks among chinese elites," *Journal of East Asian Studies*, vol. 16, no. 1, p. 1741, 2016.

4 Gregory, L. S., "A social network analysis of the chinese communist party's politburo," 2013. [Online]. Available: https://calhoun.nps.edu/handle/10945/38937

5 Jiang, J., "Making bureaucracy work: Patronage networks, performance incentives, and economic development in china," 2017.

6 Evans, P. B., *Embedded autonomy: States and industrial transformation.* Princeton University Press, 2012.

7 Macaulay, S., "Non-contractual relations in business: A preliminary study," in *The Law and Society Canon.* Routledge, 2018, pp. 155–167.

8 Gates, G., "The wen family empire," 2012. [Online]. Available: https://archive.nytimes.com/www.nytimes.com/interactive/2012/10/25/business/the-wen-family-empire.html

9 Miller, B. S., *The Powers of Art: Patronage in Indian Culture.* Oxford University Press, 1992.

10 Otte, E. and Rousseau, R., "Social network analysis: a powerful strategy, also for the information sciences," *Journal of Information Science*, vol. 28, no. 6, pp. 441–453, 2002. [Online]. Available: https://doi.org/10.1177/016555150202800601

11 Scott, J., "Social network analysis," *Sociology*, vol. 22, no. 1, pp. 109–127, 1988. [Online]. Available: https://doi.org/10.1177/0038038588022001007

12 Debnath, S., Ganguly, N., and Mitra, P., "Feature weighting in content based recommendation system using social network analysis," in *Proceedings of the 17th International Conference on World Wide Web*, ser. WWW '08. New York, NY, USA: ACM, 2008, pp. 1041–1042. [Online]. Available: http://doi.acm.org/10.1145/1367497.1367646

13 Hummon, N. P. and Dereian, P., "Connectivity in a citation network: The development of dna theory," *Social Networks*, vol. 11, no. 1, pp. 39 – 63, 1989. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0378873389900178

14 Bauer-Mehren, A., Bundschus, M., Rautschka, M., Mayer, M. A., Sanz, F., and Furlong, L. I., "Gene-disease network analysis reveals functional modules in mendelian, complex and environmental diseases," *PLOS ONE*, vol. 6, no. 6, pp. 1–13, 06 2011. [Online]. Available: https://doi.org/10.1371/journal.pone.0020284

15 Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B., "Measurement and analysis of online social networks," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '07. New York, NY, USA: ACM, 2007, pp. 29–42. [Online]. Available: http://doi.acm.org/10.1145/1298306.1298311

16 Liben-Nowell, D. and Kleinberg, J., "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20591

17 Girvan, M. and Newman, M. E. J., "Community structure in social and biological networks," *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002. [Online]. Available: http://www.pnas.org/content/99/12/7821

18 You, J., Ying, R., Ren, X., Hamilton, W., and Leskovec, J., "GraphRNN: Generating realistic graphs with deep auto-regressive models," in *Proceedings of the 35th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, Dy, J. and Krause, A., Eds., vol. 80. Stockholmsmssan, Stockholm Sweden: PMLR, 10–15 Jul 2018, pp. 5708–5717. [Online]. Available: http://proceedings.mlr.press/v80/you18a.html

19 Seibert, S. E., Kraimer, M. L., and Crant, J. M., "What do proactive people do? a longitudinal model linking proactive personality and career success," *Personnel psychology*, vol. 54, no. 4, pp. 845–874, 2001.

20 Haraway, D. J., "gender for a marxist dictionary: the sexual politics of a word," in *Women, Gender, Religion: A Reader*. Springer, 2001, pp. 49–75.

21 Blau, F. D. and DeVaro, J., "New evidence on gender differences in promotion rates: An empirical analysis of a sample of new hires," *Industrial Relations: A Journal of Economy and Society*, vol. 46, no. 3, pp. 511–550, 2007.

22 Hall, J. A., *Nonverbal sex differences: Accuracy of communication and expressive style.* Johns Hopkins University Press, 1990.

23 Remland, M. S., Jones, T. S., and Brinkman, H., "Interpersonal distance, body orientation, and touch: Effects of culture, gender, and age," *The Journal of social psychology*, vol. 135, no. 3, pp. 281–297, 1995.

24 Grover, A. and Leskovec, J., "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016, pp. 855–864.

25 Kleinberg, J. M., "Hubs, authorities, and communities," *ACM Comput. Surv.*, vol. 31, no. 4es, Dec. 1999. [Online]. Available: http://doi.acm.org/10.1145/345966.345982

26 Data, C. N., "Gdp 2017 is quarterly data (see china national data)," 2017. [Online]. Available: http://data.stats.gov.cn/english/

Appendix A: Level Annotation and Corresponding U.S. Positions

Level	Annotation	Annotation Translated	Corresponding Position-Level in U.S. (appx.)
0	无级别	No Level	n/a
1	小于副处	Below-Deputy-Division-Head level	n/a
2	副处	Deputy-Division-Head level	n/a
3	正处	Division-Head level	n/a
4	副厅	Deputy-Bureau-Director level	Municipality Mayors
5	正厅	Bureau-Director level	State Treasurers / State Controllers / County Executives
6	副部	Sub-Provincial (Ministerial) level	Lieutenant Governors / Secretaries of State
7	正部	Provincial-Ministerial level	General Cabinet Officials / State Governors
8	副国	Sub-national leader	U.S. Vice President / U.S. Secretary of State
9	正国	National leader	U.S. President

Appendix B: Regression Table for Predicting Career Ranks

	(1) Null Model	(2) OLS Model	(3) Logit Model
	Dependent variable:Cadre Final Rank		
	(1)	(2)	(3)
Constant	5.784*** (0.015)	5.071*** (0.122)	5.071*** (0.122)
V2		0.009** (0.004)	0.009** (0.004)
V3		0.0001*** (0.00004)	0.0001*** (0.00004)
V4		0.0001*** (0.00002)	0.0001*** (0.00002)
V5		0.004 (0.002)	0.004 (0.002)
V6		−0.0001 (0.00004)	−0.0001 (0.00004)
V7		−0.00003* (0.00002)	−0.00003* (0.00002)
V8			
V9		−0.00000*** (0.00000)	−0.00000*** (0.00000)
V10		−0.00000*** (0.00000)	−0.00000*** (0.00000)
V11			
V12			
V13			
V14		−0.011*** (0.004)	−0.011*** (0.004)
V15		0.00004*** (0.00001)	0.00004*** (0.00001)
V16		0.00004*** (0.00001)	0.00004*** (0.00001)
V17			
V18		0.00000 (0.00000)	0.00000 (0.00000)
V19		0.00000 (0.00000)	0.00000 (0.00000)
V20			
V21			
V22			
V23		0.00003 (0.00005)	0.00003 (0.00005)
V24		0.00000*** (0.00000)	0.00000*** (0.00000)
V25		−0.00000*** (0.00000)	−0.00000*** (0.00000)
V26			
V27		−0.000*** (0.000)	−0.000*** (0.000)
V28		0.000*** (0.000)	0.000*** (0.000)
Observations	3,021	3,021	3,021
R^2	0.000	0.407	
Adjusted R^2	0.000	0.403	
Log Likelihood			-2,976.254
Akaike Inf. Crit.			5,990.508
Residual Std. Error	0.841 (df = 3020)	0.650 (df = 3002)	
F Statistic		114.414*** (df = 18; 3002)	

Note: *p<0.1; **p<0.05; ***p<0.01