Complex Network Analysis of Ozone Transport

Guoxun Tian and Mehmet Hadi Gunes

Abstract. Ozone transport has an important effect on the local ozone concentrations. In this paper, we model the California State and the Eastern States of America ozone transport as complex networks to understand large-scale dynamics. California has a complex ozone transport due to geomorphology and is divided into 15 air basins to manage air pollution. Unlike California, the ozone transport in the Eastern States of America is mainly affected by wind speed and direction. Through different centralities, we can identify nodes that have higher ozone output, higher ozone income, and the maximum ozone throughput. In general, both networks exhibit similar properties. However, even though Californian network has higher average degree it has smaller clustering than the Eastern State network. Moreover, while Californian counties can be divided into four communities, the Eastern States remain as a single community. Both of these points to the fact that ozone transport within Eastern States is more uniform than between the counties of California.

1 Introduction

In the upper atmosphere, ozone protects the earth from exposure to harmful ultraviolet rays since ozone has a strong absorption of ultraviolet [1]. However, at the ground level, ozone is one of the most important air pollutants. The harms of human health are considerable [2]. In this paper, we establish an *ozone transport network* to analyze large scale characteristics of ozone transfer. To our knowledge, this is the first study to map the ozone transport as a complex network.

Control of ozone emissions is not very effective in reducing ozone concentrations since there are relatively few anthropogenic sources of ozone. Ozone is

Guoxun Tian
Colorado State University, Fort Collins, CO, USA
e-mail: tgxfff@hotmail.com

Mehmet Hadi Gunes
University of Nevada, Reno, NV, USA
e-mail: mgunes@cse.unr.edu

P. Contucci et al. (eds.), *Complex Networks V*,
Studies in Computational Intelligence 549,
DOI: 10.1007/978-3-319-05401-8_9,

primarily formed in the air. After the formation, ozone is cumulated in the air at day time with sunlight and is transported to other places, propagating the air pollution, affecting the people in a downwind area which may be far away from the pollution sources. Transport and formation are the two main sources of ozone pollution.

Long-range transport of ozone and its precursors has become a question in the West since they are isolated to a local area because of the topography. The United States Environmental Protection Agency (US EPA) use a new ozone standard based on the average of ozone concentrations. Areas failing to meet these standards set by the EPA were defined as a “non-attainment areas” for ozone [3].

There were 4 states in the West with areas that do not comply with new standard, including California, Nevada, Arizona, and Colorado. Among which California was the worst. Several factors have to be taken into account to figure out why California has the highest ozone level and so many non-attainment areas. The first one is the long distance pollutant transport from Eastern Asia. Also, the ozone and its precursor can be transported from Eastern Asia to America. Several studies show that the concentration of air pollutants in Eastern Asia significantly affects the background levels of ozone in Western North America [4]. The second factor we have to consider is the ozone transport between different areas in California. To better manage air ozone transport, California is divided into 15 air basins based on its topography features or political boundary [5]. There are 68 counties are included in these 15 air basins. In this study, a complex network is built for these 68 counties to analyze the transport of ozone between them so the responsibility can be better understood.

In the Eastern States of America, control programs throughout the northeast and southeast was established due to the formation and transport of ozone over long distances. These control programs are often more regional in scope. Through these programs, Eastern States and local governments have found that pollutants which are transported from the large industrial combustion sources in the Midwest and Southeast contribute to non-attainment classifications in these areas. The speed and direction of wind is the main factor of this long-range transport. It was found that moderate to high wind speeds have a moderate to high potential for a contribution from transport [6]. That is high pollution episodes from the Midwest and Southeast, which has heavy NOx pollution, can be taken by airflow streams to the Northeast. This long distance transport has been displayed using back trajectory analysis. In this study, the complex network of ozone transport between Eastern States is also constructed to better understand the ozone transport.

2 Californian County Network

2.1 Ozone Transport Classification between Air Basins

California State is divided to 15 air basins for the ozone transport research. In this paper, we divide it into 68 counties for a finer granularity analysis using complex network metrics. In our network, the nodes represent counties and the edges represent the ozone transport between them. Each county is treated as a sub-region of the corresponding air basin. The transport between air basins is estimated based on the assessment of the impacts of transported pollutants on ozone concentrations.

California Environmental Protection Agency Air Resources Board (CEPAARB) estimate the transported pollutants on ozone concentrations, identify upwind air basins and the downwind air basins, and make an assessment of the relative contributions of upwind emissions to downwind ozone concentrations [7]. Classification of aerosol transport between different regions made by CEPAARB is [5]: *Inconsequential*, *Significant*, and *Overwhelmed*.

Using statistical analysis, we can estimate the transport effect by comparing the diurnal profile of hourly ozone concentration averaged over all potential transport days with similar plots for days not in the potential transport [8]. In this network, the weight of edges is determined based on the ozone transport classification between air basins using data from [7]. We assign weights of 0.1 to *inconsequential*, 0.6 to *significant*, and 1 to *overwhelmed*. The average value of the combinations represents the weight of the edges. As [7] only shows the ozone transport between air basins, refinement of the air basins is needed. Thereby, we estimate the ozone transport between different counties using following assumptions.

- **Assumption 1:** Ozone transport only exists between two adjacent counties. If these two counties are at different air basins, the edge between them has the same weight as the ozone transport between the two adjacent air basins.

This assumption is used to estimate the ozone transport between two counties which are included in different air basins. For example, air transport from San Joaquin Valley to North Central Coast has weight of 0.6. Fresno County is in San Joaquin Valley and adjacent to Mono County which is in North Central Coast. So the ozone transport from Fresno County to Mono County also has weight 0.6.

- **Assumption 2:** Air pollution can move freely within an air basin. In the same air basin, the maximum transport weight is set to 1. The actual weight value is directly proportional to the NOx emission of the air basin since it plays an important role in the formation of ozone. The quantification rule is:

  a) Within the air basin which has the maximum NOx emission (MNE) which is equal to 34% of state total, the ozone transport has a weight of 1.
  b) Within other air basins, the weight of the ozone transport is directly proportional to the air basin NOx emission (NE) using the equation $Weight = \frac{NE}{MNE}$.

For example, in San Francisco Bay Area air basin, there is a heavy concentration of industrial facilities, several airports, and a dense freeway and surface street network. The NOx emission has 16% of state total. San Francisco County and San Mateo County are adjacent, so the weight between two of them is quantified as 0.16/0.34=0.47. However, in North Coast air basin, the northern part is sparsely populated and the air is very clean. So the ozone transport in that area is neglected.

- **Assumption 3:** Ozone transport between non-adjacent counties is ignored.

For example, ozone transport from San Francisco County to Mountain Counties and South Central Coast is neglected since they are not adjacent. In reality, ozone transport does exist between non-adjacent counties. It is neglected in this study because the ozone transport through aloft flow is much less than the direct wind transport, and each county always has a lot of non-adjacent neighbors.

2.2 Analysis of the Californian County Network

In this section, we analyze different measures of the ozone transport network of Californian counties.

Closeness Centrality: In a graph, the farness of a node is defined as the sum of its distances to all other nodes, and its closeness is defined as the inverse of the farness, i.e., $closeness(i) = \frac{1}{\sum_j d_{ij}}$ where i is the focal node, j is any node in the network and d_{ij} is the shortest distance between them. As a consequence, a node with higher closeness has lower total distance to all other nodes. For a network, closeness is used to estimate how long it will take to spread information from one node to other nodes sequentially.

In our case, instead of information spreading, ozone transport is considered. *A node that has higher closeness can transport ozone to other nodes in a shorter time*. The closeness centrality value give us a hint about which nodes has the higher ozone output. The average closeness is 0.164 and the top 10 counties have closeness centrality higher or equal to 0.237.

Compared to other air basins, San Francisco Bay Area Air Basin has the most counties in the top 10 in terms of centrality. This is consistent with the history since this air basin has violated the state and federal health-based standards many times over the years, and has contributed to air pollution problems in all of the surrounding air basins.

Betweenness Centrality: Betweenness centrality is a measure of a node within a network. It was introduced as a measure for quantifying the control of a node on the communication between other nodes. It is defined as $C(n) = \sum_{i,j} \frac{\sigma_{i,j}(n)}{\sigma_{i,j}}$ where $\sigma_{i,j}$ is the total number of shortest paths between nodes i and j and $\sigma_{i,j}(n)$ is the number of shortest paths between nodes i and j that pass node n. Nodes of high betweenness centrality are important since communication or transport is more efficient along the shortest paths.

For this ozone transport network, *the node with higher betweenness has the higher ozone throughput*. The average betweenness centrality of this network is 0.010 and the top 10 counties have betweenness higher or equal to 0.044. Despite the nodes with highest betweenness, compared to other air basins, San Joaquin Valley Air Basin includes the biggest portion, five counties, among the top 10 counties. This air basin is at the center of California State and has the second highest NOx emission. The location and high NOx emission give this air basin the ability to affect other areas.

PageRank: The PageRank value indicates an importance of a particular page. A hyperlink to a page counts as a vote of support. A page with high PageRank means it receives a high rank and it is linked "incoming link" to by many pages. It is defined as $PR(p_i) = \frac{1-d}{N} + d\sum_{p_j \in M(p_i)} \frac{PR(p_j)}{L(p_j)}$ where N is the total number of pages, $p_1, p_2, \ldots\ldots, p_N$ are the pages under consideration, $L(p_j)$ is the number of out bound links on page p_j, and $M(p_i)$ is the set of pages that link to p_i.

In this application, instead of a page, a node of the ozone transport is considered. *A node that has higher PageRank is easier to be contaminated by incoming ozone transport from other nodes*. The average PageRank of this network is 0.011 and the top 10 counties have PageRank greater or equal to 0.026. Mojave Desert Air Basin, which is heavily impacted by other air basins, includes three of them.

Clustering: Clustering coefficient has been defined as $\mathrm{C} = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of nodes}}$ or $C = \frac{\textit{number of closed triplets}}{\textit{number of connected triples of nodes}}$. The average clustering coefficients of the Californian County network is 0.25. The clustering coefficients are shown in Figure 1.

Assortativity: Assortativity coefficient indicates how nodes of different types are connected amongst themselves, and is defined by $\mathrm{r} = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$ where $a_i = \sum_j e_{ij}$ and $b_j = \sum_i e_{ij}$ and e_{ij} is the fraction of edges from node of type i to a node of type j. Assortativity coefficient of the ozone transport network is 0.18, which means this network is slightly assortative, i.e., nodes of high or low degrees are a bit likely to affect similar degree areas.

Average Degree: The average degree of this network is 5.06. Degree distribution in Figure indicates that the network has an exponential degree distribution. Different from typical complex networks that have power law degree distribution, each node has approximately the same number of nodes connected to them.

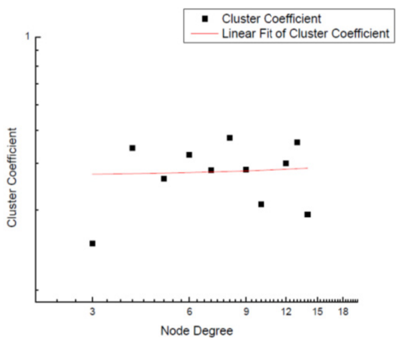

Fig. 1 Clustering coefficients distribution of Californian County network (0 clustering and 0 degree nodes are ignored)

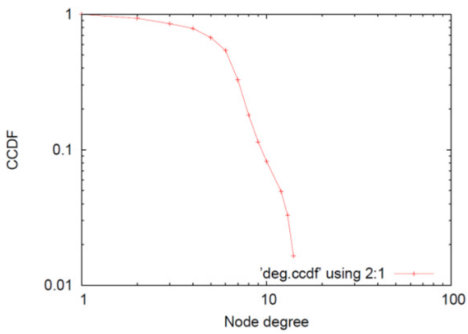

Fig. 2 Degree distribution of Californian County network (0 degree nodes are ignored)

Community Structure: Community detection mechanisms try to identify strongly connected communities within a given network. We use simulated annealing approach which tries to minimize the Hamiltonian $H(\{\sigma\}) = -\sum_{i\neq j}(A_{ij} - \gamma p_{ij})\,(\sigma_i, \sigma_j)$ where p_{ij} is the probability of vertices i and j being connected. It can be shown that minimizing this Hamiltonian, with $\gamma = 1$ is equivalent to maximizing Newman's modularity. By increasing the parameter γ, it's possible also to find sub-communities. As shown in Figure 3, the network can be divided to 4 groups, i.e., network consists of four regions that are more densely interrelated.

Fig. 3 Communities of the Californian ozone transport network. Different color represents different group.

3 Eastern State Network

In this section, we establish a non-weighted Eastern State network. Transport of ozone and precursors between the Eastern States of America has no boundaries and it is highly related to the wind speed and direction. Ozone can travel across states and provinces easily. At different wind directions and speeds, the ozone concentration pattern is consistent with an atmospheric ozone lifetime of about one day. High ozone concentrations are typically located downwind of areas with the highest emissions with high wind speed (i.e., >6m/s).

Transport during high and low ozone days is investigated in [9]. Transport conditions were established for regionally high (90%) and low (10%) daily maximum 1-hour ozone concentrations. Since the absolute ozone transport between different states is difficult to quantify, a non-weighted ozone transport network is constructed based on the wind directions. In the graph, if wind blow air from state A to B, a non-weighted edge from A to B in the transport network is established. For example, as shown in Figure 5, wind vector shows that wind blow air from Texas to Oklahoma, so an edge from Texas (nodes 0) to Oklahoma (nodes 1) is established.

3.1 Analysis of the Eastern State Network

In this section, we analyze some network metrics of the ozone transport network of the Eastern States.

Closeness Centrality: Compared to the Californian County network, the ozone transport network of the Eastern States of America has a higher average closeness centrality. The average value is 0.278 and Illinois and Kentucky have highest closeness with value 0.367.

Betweenness Centrality: The average betweenness value of this network is 0.025, which is close to the Californian County network. New York has highest betweenness with value 0.13.

PageRank: The average PageRank of this network is 0.014 and is little higher than the Californian County network. New York State has highest PageRank with value 0.45.

Clustering: The average clustering coefficient of the ozone transport network is 0.33, which is higher than the Californian County network. This indicates higher dependence between different regions.

Assortativity: Assortativity coefficient of this ozone transport network is 0.26, which is higher than the Californian County network and shows the network is a bit more assortative. That is states of higher degree are more likely to affect or be affected by other high degree regions.

Average Degree: The average degree of this network is 3.6, which is lower than the Californian County network. Degree distribution is shown in Figure 4. Similar to Californian County network, this figure also shows that the network has an exponential degree distribution.

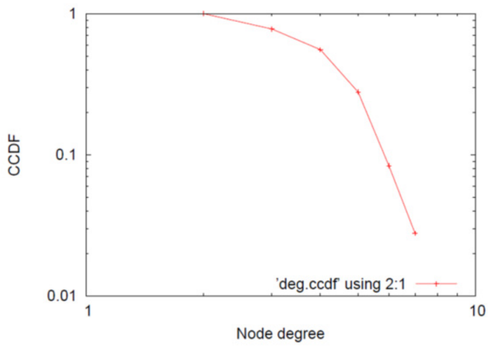

Fig. 4 Degree distribution of the ozone transport of Eastern States

Community Structure: The community structure is shown in Figure 5 where the network is divided into 4 groups. Compared with Californian County network, this network has higher average clustering coefficient, which means this network is more connected. The group barrier, however, is not clear as three of the communities can be accepted as a single group for a total of two groups.

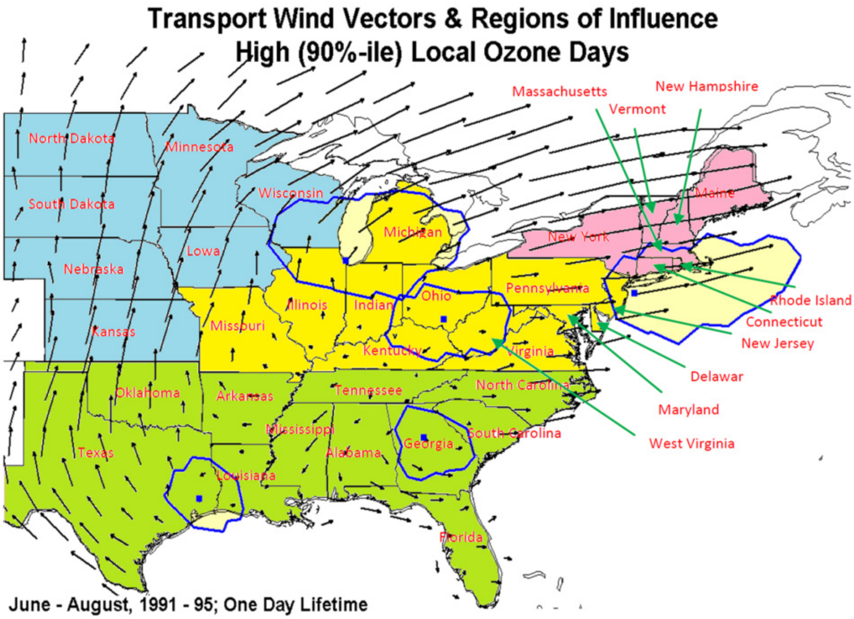

Fig. 5 The communities of ozone transport in Eastern States (modified from [9])

4 Discussion

This paper presents a new way to analyze the ozone transport. We explore the properties of the ozone transport from the complex network perspective. We establish two ozone transport networks, one for Californian Counties and another one for the Eastern States of the United States of America. Some complex network properties of these networks are different as shown in Table 1. Compared to the Californian County network, the Eastern State network has a higher average closeness centrality, a higher clustering coefficient, a higher assortativity coefficient, and a lower average degree. The PageRank centrality, betweenness centrality and degree distribution are similar. Both networks have exponential degree distribution since the ozone transport between non-adjacent areas is ignored.

Generally, the explored properties of the networks are consistent with the existing research. For example, San Francisco Bay Area Air Basin, which is heavy polluted and has contributed to air pollution problems in all of the surrounding air basins, has the biggest portion among the top 10 counties in terms of closeness centrality. Similarly, in terms of closeness centrality, 3 of the top 10 counties are included in San Francisco Bay Area Air Basin. The other 7 counties, which are distributed in several air basins, give us some details of the ozone transport. Those counties may also be playing or will potentially play important roles in the ozone transport.

Table 1 Properties of Californian County network and Eastern State network

	Californian Counties	**Eastern States**
Average closeness centrality	0.164	0.278
Average betweenness centrality	0.010	0.025
Average PageRank	0.013	0.014
Average clustering coefficients	0.25	0.33
Assortativity coefficient	0.18	0.26
Average degree	5.06	3.6
Degree distribution	Exponential	Exponential
Community structure	4 groups	2 or 4 groups

Among the Eastern States of America, Michigan, one of the biggest source areas of ozone transport plays an important role in the ozone transport network. It has the 3rd highest closeness centrality, 2nd highest betweenness centrality and 2nd highest PageRank. New York, which has the highest closeness centrality and PageRank, may be playing or will potentially play an important role in the ozone transport as well. The community structure of Eastern State network also shows that New York is a key node is the only conjunction node of the two groups.

As we can see from Table 1, these two networks have similar properties. Despite the fact that the average degree of Californian county network is higher than Eastern State network, the average clustering coefficient of Californian county network is lower since the geographic characteristics of the Californian counties. These facts point to the geomorphology effect for ozone transport and the fact that ozone transport within Eastern States is much more uniform than between the Californian counties.

Acknowledgement: This material is based in part upon work supported by the National Science Foundation under grant number EPS- IIA-1301726

Disclaimer: Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Seinfeld, J.H., Pandis, S.N.: Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. John Wiley and Sons, Inc. (1998)
[2] Avol, E.L., Linn, W.S., Venet, T.G., Shamoo, D.A., Hackney, J.D.: Comparative respiratory effects of ozone and ambient oxidant pollution exposure during heavy exercise. J. Air Pollut. Control Assoc. 34, 804–809 (1984)

[3] Wierman, S.S.: Will the 8-hr ozone standard finally get off the ground. EM, 20–28 (September 2003)
[4] Schwarzhoff, P.: Long-range Transport of Pollutants to the West Coast of North America. Georgia Basin/Puget Sound Research Conference (2003)
[5] Garcia, C., Gouze, S., Wright, W.: Assessment of the Impacts of Transported Pollutants on Ozone Concentrations in California. California Air Resources Board (2001)
[6] Douglas, S.G., Hudischewskyj, A.B., Lolk, N.K., Guo, Z.: Analysis of Southern California Wind Profiler and Aircraft Data. SAI Report (1997)
[7] Austin, J., Gouze, S.: Ozone Transport: 2001 Review. California Environmental Protection Agency Air Resources Board (2001)
[8] Tian, G.: Complex Network Analysis of Ozone Transport. MS Thesis, University of Nevada, Reno (2012)
[9] Schichtel, B.A., Husar, R.B.: Eastern north america transprot climatology during average, high and low ozone days (1999)