Institute of Smart Systems and Artificial Intelligence

NAZARBAYEV UNIVERSITY

COVID-19 Simulator for Kazakhstan

Prof. H. Atakan Varol

Founding Director of ISSAI

ahvarol@nu.edu.kz, issai.nu.edu.kz

Team: Askat Kuzdeuov, Daulet Baimukashev, Mukhamet Nurpeiissov, Aknur Karabay, Bauyrzhan Ibragimov, and Almas Mirzakhmetov

Background

The COVID-19 pandemic has emerged as a global crisis that threatens to overwhelm public healthcare systems and disrupt the social and economic welfare of every country.

The daily lives of billions are impacted by the unprecedented attempts by governments to control the spread through the imposition of strict quarantines and social distancing measures.

COVID-19 has propagated rapidly amongst a globally susceptible population, with a high mortality rate amongst victims, and the unfortunate characteristic that carriers are highly contagious in pre-symptomatic states.

The situation is worsened by the fact that the population has no prior exposure, and there is no vaccine available, and that widespread testing for the virus began quite late due to lack of testing kits and facilities.

In this scenario, epidemiological models can be used to project the future course of the disease, and to estimate the impact of measures that might be used to control its spread.

Project Objectives

- To help in this time of crisis, the ISSAI team of Nazarbayev University has developed a stochastic epidemic simulator, calibrated with real-world experience, and customized for the Republic of Kazakhstan.

- The simulator utilizes real RoK data, ranging from population density to health care capacity for each region to predict the dynamics of the spread of COVID-19 in Kazakhstan, and thereby inform government policy-making.

- We model Kazakhstan as a graph of 17 connected nodes (14 oblasts and 3 cities of Republican significance) where each node runs a separate SEIR epidemiological model.

- Based on air, rail and highway connections between the nodes, we also model the spread of the disease between different geographic regions.

The RoK COVID-19 Simulator

- The simulator is initially configured to model the current situation and project the ongoing impact, which can be used to facilitate government allocation of resources.

- The real value of the simulator is that it enables the modeling of disease suppression and mitigation measures, such as limiting travel or quarantining a region, on both a localized and national basis.

- The simulator can be used to estimate the impact and duration of the epidemic over time, based on a range of policy measures, and thus allow better state planning to reduce the impact.

- ISSAI has shared the source code such that it can be used, adapted, and improved by others.

https://github.com/baimukashev/COVID-19_simulation/tree/master

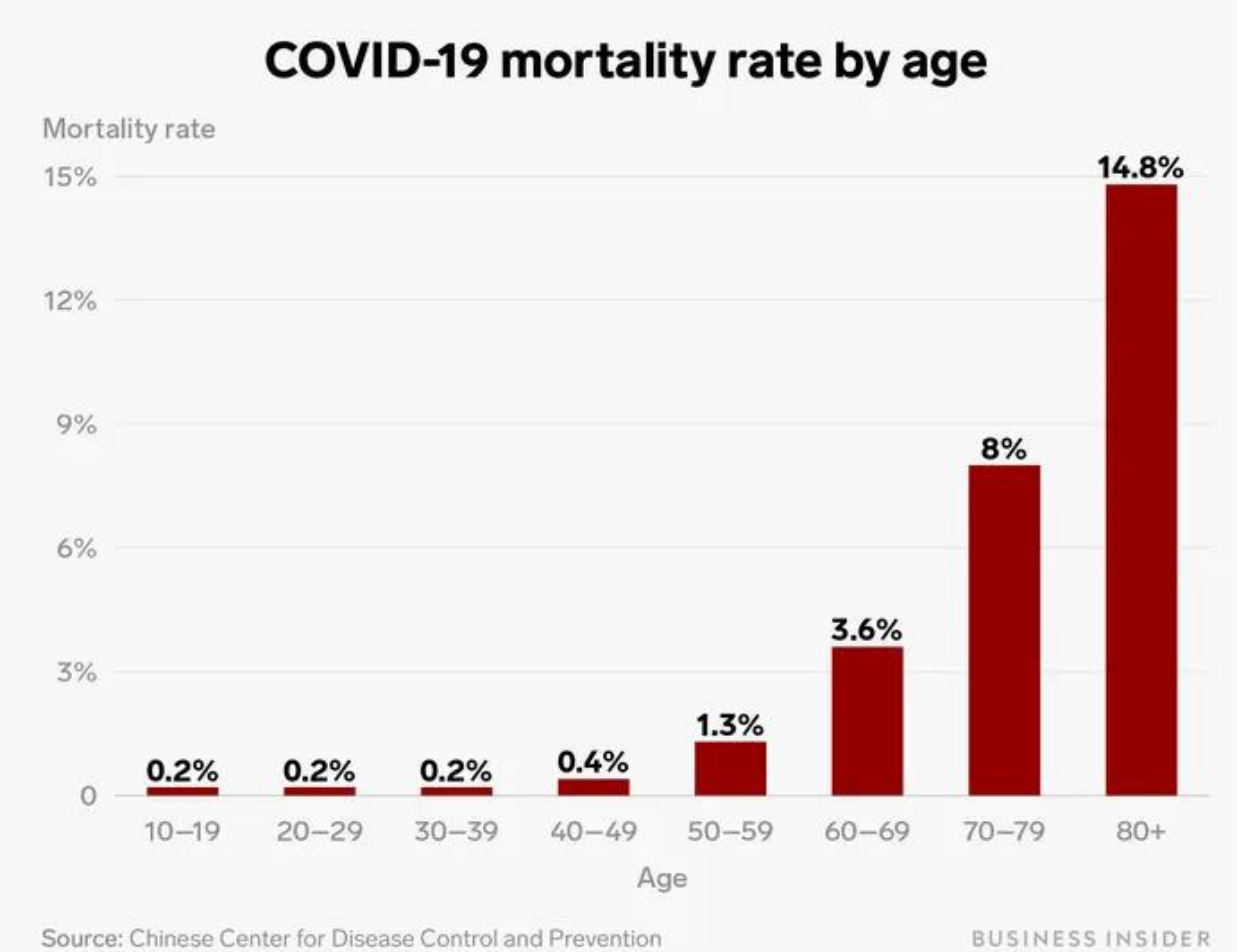
Disclaimer: This is a research tool which will show general future trends based on the entered parameters. Projected outcomes are dependent on correct initial conditions, and the accuracy of the parameters.

Part I – Underlying Assumptions:

Incorporating the experience of China, Diamond Princess, and Lombardy

Reported Mortality of the Confirmed COVID-19 Cases

- In China, the overall mortality rate is reported as 3.8 percent.
- Mortality is high for 70+, even higher for 80+.



COVID-19 Epidemic in Diamond Princess Cruise Ship

- The COVID-19 epidemic on the Diamond Princess cruise ship provides valuable information since it was a closed system conducive to disease transmission with lots of testing conducted on the passengers.
- One third of passengers are 70+.
- Out of 3711 passengers, 619 got infected in 3 weeks hinting rapid disease transmission.
- 7 deaths are reported out of 619 cases resulting in a mortality rate of 1.1%.
- According to the China COVID-19 mortality rates, 35 deaths would be expected.

Age Range	No. of passengers	Symp. cases	Asymp. cases	nCFR	Expected deaths using nCFR	Observed deaths on cruise ship
0 - 9	16	0	1	0.0% (0.0% - 0.9%)	0 (0 - 0)	0
10 - 19	23	2	3	0.2% (0.0% - 1.0%)	0 (0 - 0)	0
20 - 29	347	25	3	0.2% (0.1% - 0.4%)	0.05 (0.02 - 0.10)	0
30 - 39	428	27	7	0.2% (0.1% - 0.4%)	0.06 (0.04 - 0.10)	0
40 - 49	334	19	8	0.4% (0.3% - 0.6%)	0.08 (0.06 - 0.12)	0
50 - 59	398	28	31	1.3% (1.1% - 1.5%)	0.36 (0.31 - 0.43)	0
60 - 69	923	76	101	3.6% (3.2% - 4.0%)	2.74 (2.5 - 3.1)	0
70 - 79	1015	95	139	8.0% (7.2% - 8.9%)	7.6 (6.8 - 8.4)	6
80 - 89	216	29	25	14.8% (13.0% - 16.7%)	4.28 (3.8 - 4.9)	1
Totals	3711	301	318		15.15 (13.5 - 17.1)	7

https://cmmid.github.io/topics/covid19/severity/diamond_cruise_cfr_estimates.html

Population Pyramids and Expected Mortality Rates

- Since mortality rate is higher for elderly and the average age in Diamond Princess was 57, the COVID-19 related mortality rate in Diamond Princess should have been on the higher end of the scale.

- Using the age-related mortality ratios from the China data, 1.1% mortality rate of Diamond Princess, and population pyramids, we expect the mortality rates of **Lombardy** (epicenter of COVID-19 in Italy) and **Kazakhstan** to be **0.7%** and **0.3%**, respectively.

Ratio of Population in Different Age Groups

Age Group	Diamond Princess	Lombardy	Kazakhstan
0 - 9	0.004	0.088	0.120
10 - 19	0.006	0.096	0.127
20 - 29	0.094	0.098	0.068
30 - 39	0.115	0.118	0.215
40 - 49	0.071	0.158	0.210
50 - 59	0.107	0.156	0.112
60 - 69	0.249	0.118	0.080
70 - 79	0.274	0.099	0.061
80+	0.058	0.077	0.007

Note that Kazakhstan has much lower percentage of persons aged 70+ and especially 80+.

How to Test Assumptions on the Mortality Rate?

- **Large Scale Testing:** A large portion of the population can be tested randomly for COVID-19. This would provide reliable estimates on the number of infected with COVID-19. Based on the number of deaths, the mortality rate can be computed. → **not very practical.**

- **Large Scale Serological Testing:** Serological tests looks for antibodies in blood. They can identify people who were infected and recovered from COVID-19. Combined with random sampling, the full scope of the epidemic can be estimated. As of now, **there is no approved serological test for COVID-19.**

- In order to test our assumptions, we will first focus on the simulation of Lombardy region in Italy.
- Lombardy is the epicenter of COVID-19 in Italy.
- Its population is around 10 million.
- There are already over 6,000 COVID-19 related deaths.
- The epidemic timeline is well established.
- Italian government shares its data daily since 24 February 2020.

(https://github.com/pcm-dpc/COVID-19/tree/master/dati-regioni)

A Closer Look to the Lombardy COVID-19 Data

Day	Date	Hospitalized with Symptoms	Intensive Care	Total Hospitalized	Home Isolation	Current Cases	New Cases	Recovered	Dead	Total Cases	Tests
0	2/24/2020	76	19	95	71	166	166	0	6	172	1463
1	2/25/2020	79	25	104	127	231	65	0	9	240	3700
2	2/26/2020	79	25	104	145	249	18	0	9	258	3208
3	2/27/2020	172	41	213	136	349	100	40	14	403	3320
4	2/28/2020	235	47	282	192	474	125	40	17	531	4835
5	2/29/2020	256	80	336	216	552	78	40	23	615	5723
6	3/1/2020	406	106	512	375	887	335	73	24	984	6879
7	3/2/2020	478	127	605	472	1077	190	139	38	1254	7925
8	3/3/2020	698	167	865	461	1326	249	139	55	1520	9577
9	3/4/2020	877	209	1086	411	1497	171	250	73	1820	12138
10	3/5/2020	1169	244	1413	364	1777	280	376	98	2251	12354
11	3/6/2020	1622	309	1931	77	2008	231	469	135	2612	13556
12	3/7/2020	1661	359	2020	722	2742	734	524	154	3420	15778
13	3/8/2020	2217	399	2616	756	3372	630	550	267	4189	18534
14	3/9/2020	2802	440	3242	1248	4490	1118	646	333	5469	20135
15	3/10/2020	3319	466	3785	642	4427	-63	896	468	5791	21479
16	3/11/2020	3852	560	4412	1351	5763	1336	900	617	7280	25629
17	3/12/2020	4247	605	4852	2044	6896	1133	1085	744	8725	29534
18	3/13/2020	4435	650	5085	2647	7732	836	1198	890	9820	32700
19	3/14/2020	4898	732	5630	3429	9059	1327	1660	966	11685	37138
20	3/15/2020	5500	767	6267	3776	10043	984	2011	1218	13272	40369
21	3/16/2020	6171	823	6994	3867	10861	818	2368	1420	14649	43565
22	3/17/2020	6953	879	7832	4263	12095	1234	2485	1640	16220	46449
23	3/18/2020	7285	924	8209	4057	12266	171	3488	1959	17713	48983
24	3/19/2020	7387	1006	8393	5545	13938	1672	3778	2168	19884	52244
25	3/20/2020	7735	1050	8785	6635	15420	1482	4295	2549	22264	57174
26	3/21/2020	8258	1093	9351	8019	17370	1950	5050	3095	25515	66730
27	3/22/2020	9439	1142	10581	7304	17885	515	5865	3456	27206	70598
28	3/23/2020	9266	1183	10449	8461	18910	1025	6075	3776	28761	73242
29	3/24/2020	9711	1194	10905	8963	19868	958	6657	4178	30703	76695
30	3/25/2020	10026	1236	11262	9329	20591	723	7281	4474	32346	81666
31	3/26/2020	10681	1263	11944	10245	22189	1598	7839	4861	34889	87713
32	3/27/2020	11137	1292	12429	11466	23895	1706	8001	5402	37298	95860
33	3/28/2020	11152	1319	12471	12038	24509	614	8962	5944	39415	102503

A Closer Look to the Lombardy COVID-19 Data

Day	Date	Hospitalized with Symptoms	Intensive Care	Total Hospitalized	Home Isolation	Current Cases	New Cases	Recovered	Dead	Total Cases	Tests
0	2/24/2020	76	19	95	71	166	166	0	6	172	1463
1	2/25/2020	79	25	104	127	231	65	0	9	240	3700
2	2/26/2020	79	25	104	145	249	18	0	9	258	3208
3	2/27/2020	172	41	213	136	349	100	40	14	403	3320
29	3/24/2020	9711	1194	10905	8963	19868	958	6657	4178	30703	76695
30	3/25/2020	10026	1236	11262	9329	20591	723	7281	4474	32346	81666
31	3/26/2020	10681	1263	11944	10245	22189	1598	7839	4861	34889	87713
32	3/27/2020	11137	1292	12429	11466	23895	1706	8001	5402	37298	95860
33	3/28/2020	11152	1319	12471	12038	24509	614	8962	5944	39415	102503

- Only 1 percent of the population is tested. Among tested 40% have COVID-19.
- Nearly 6000 deaths occurred with the latest daily number of 542 deaths (28 March 2020).
- In the initial days of recording, one fourth of the hospitalized patients are in intensive care units (ICUs). Now, it is 12 percent. This suggests saturation of the ICU capacity and possibly might account for part of the higher mortality rate compared to the Diamond Princes.
- 28 percent of the positive tested cases are hospitalized.

Reconnecting Lombardy and Diamond Princess

- Lombardy already has 6,000 deaths. If the mortality rate is 0.7 percent as predicted, there should be 850,000 infected in Lombardy.
- Considering that the total number of confirmed cases in the world as of 29 March 2020 is 677,000, 850,000 appears to be extreme.
- Presumably, the mortality rate in Lombardy might be higher than 0.7 percent because intensive care units (ICUs) are overloaded and cannot serve new patients.
- However, even if we account for the increased mortality due the saturated ICUs, there should be 500,000 infected in Lombardy.

- This suggests that mild or asymptomatic cases of COVID-19 are heavily underreported.

GOOD: COVID-19 is not as deadly as predicted.
BAD: There are much more infected people around. It is hard to contain this epidemic.
UGLY: We don't have enough random tests or any serological tests to verify these claims.

- What should we do?
  - 1) Look at history
  - 2) Simulate Lombardy and find a model that fits.

Past as a Good Predictor of the Future
2009 H1N1 Influenza Epidemic in the United States

CDC Estimates of 2009 H1N1 Cases and Deaths from April 2009 - February 13, 2010, By Age

2009 H1N1	Mid-Level Range*	Estimated Range*
Cases		
0-17 years	~19 million	~14 million to ~28 million
18-64 years	~34 million	~24 million to ~50 million
65 years and older	~6 million	~4 million to ~8 million
Cases Total	~59 million	~42 million to ~86 million
Hospitalizations		
0-17 years	~85,000	~60,000 to ~125,000
18-64 years	~154,000	~109,000 to ~226,000
65 years and older	~26,000	~19,000 to ~38,000
Hospitalizations Total	~265,000	~188,000 to ~389,000
Deaths		
0-17 years	~1,250	~890 to ~1,840
18-64 years	~9,200	~6,530 to ~13,500
65 years and older	~1,550	~1,100 to ~2,280
Deaths Total	~12,000	~8,520 to ~17,620

After a three-year analysis of multiple research groups, Center for Disease Control predicted **59 million cases** for H1N1 (Swine Flu), **265,000 hospitalizations,** and **12,000 deaths**.

Case Mortality Rate is 12,000 over 59,000,000 → 0.02 percent.

Our mathematical analysis predicted the mortality rate of COVID-19 as 0.7 percent (35 times H1N1) for Lombardy and 0.3 percent (15 times H1N1) for Kazakhstan.

https://www.cdc.gov/H1N1flu/estimates/April_February_13.htm

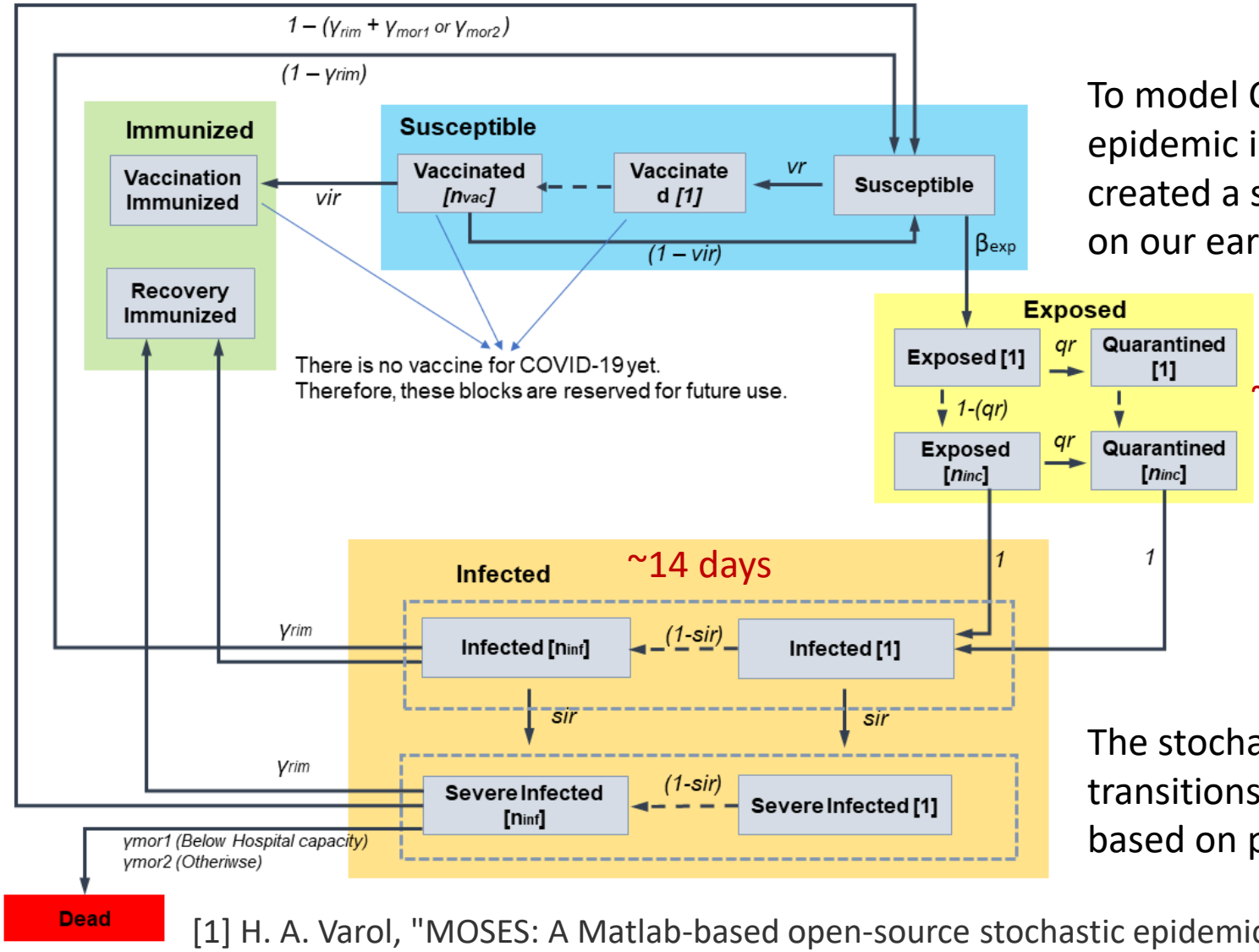
Lombardy COVID-19 Timeline

Day	Date	Death	Confirmed Cases	Tests	Event
-36	**1/21/2020**	**??**	**??**	**??**	COVID-19 infected person comes from China.
-10	**2/14/2020**	**??**	**??**	**??**	Male person who met the person from China feels bad, goes to doctor, he and his wife were later confirmed positive for COVID-19.
-4	**2/20/2020**	**??**	**3**	**??**	Three confirmed cases in Lombardy.
0	**2/24/2020**	**6**	**172**	**1463**	Data is saved regularly in the national repository. COVID-19 was already spreading for 36 days with already 6 deaths.
7	**3/2/2020**	38	1254	7925	Red zone in Lombardy locked down.
13	**3/8/2020**	267	4189	18534	Whole Lombardy locked down.
16	**3/11/2020**	617	7280	25629	All bars and restaurants are closed
27	**3/22/2020**	3456	27206	70598	Factories, nonessential production closed.
34	**3/28/2020**	5944	39415	102503	Date of last data available.

- On 24 February 2020, when there were 6 registered COVID-16 deaths in Lombardy, we assume there was already 857 cases of COVID-19 in Lombardy (based on 0.77 percent mortality rate estimated for Italy).

- Now, we need a model to simulate Lombardy region.
- If our model predicts the current recorded deaths accurately, we might either have a good model or we simply might be overfitting our model to the available data.
- If our model shows good predictive accuracy for the next one week or so, then we can state that we have a good model and our assumptions were solid.

Stochastic SEIR-S Model Simulator

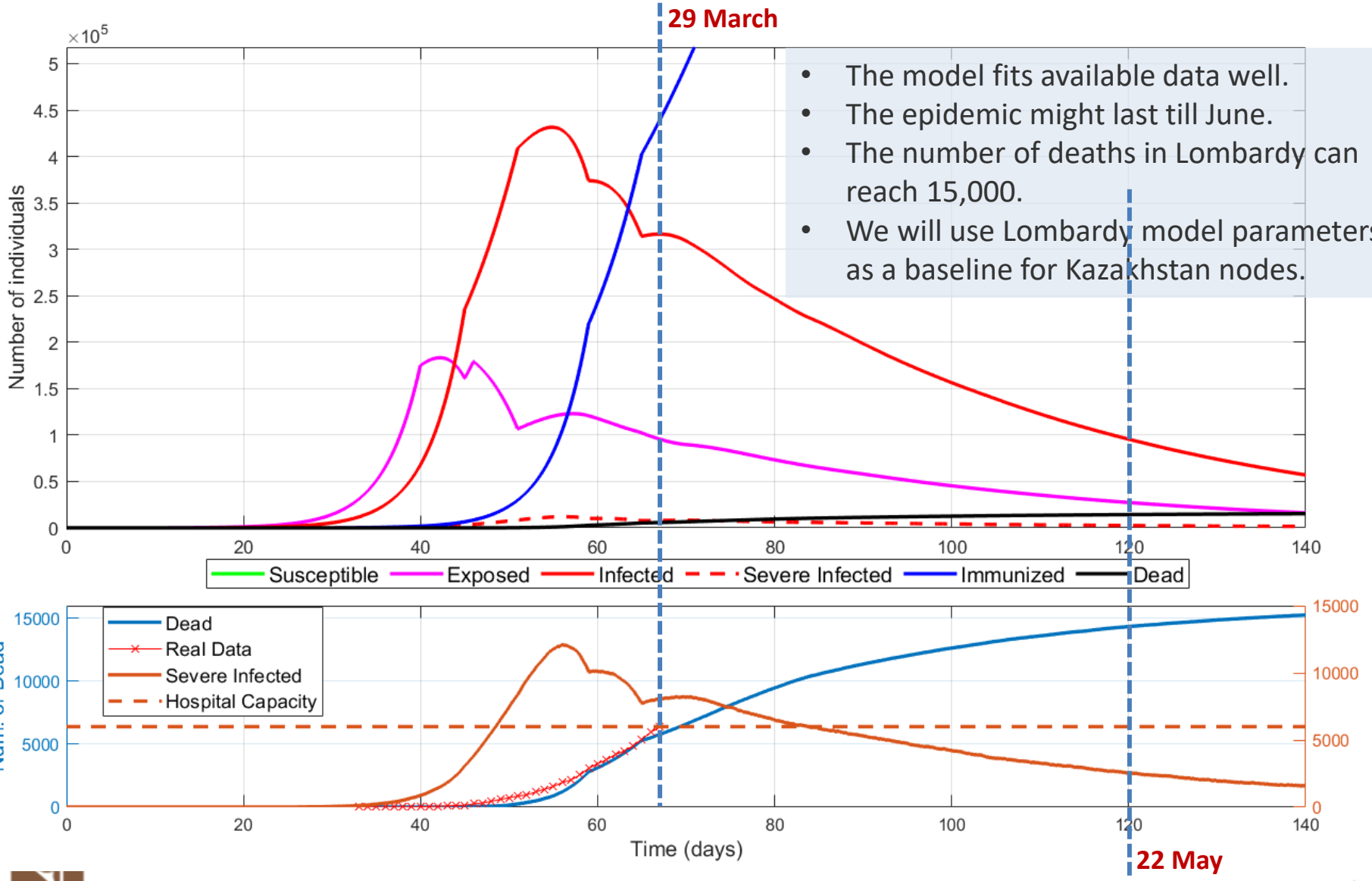


To model COVID-19 epidemic in Lombardy, we created a simulator based on our earlier work [1].

The stochastic model transitions between states based on probabilities.

[1] H. A. Varol, "MOSES: A Matlab-based open-source stochastic epidemic simulator," *IEEE Int. Conf. of the Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 2636-2639.

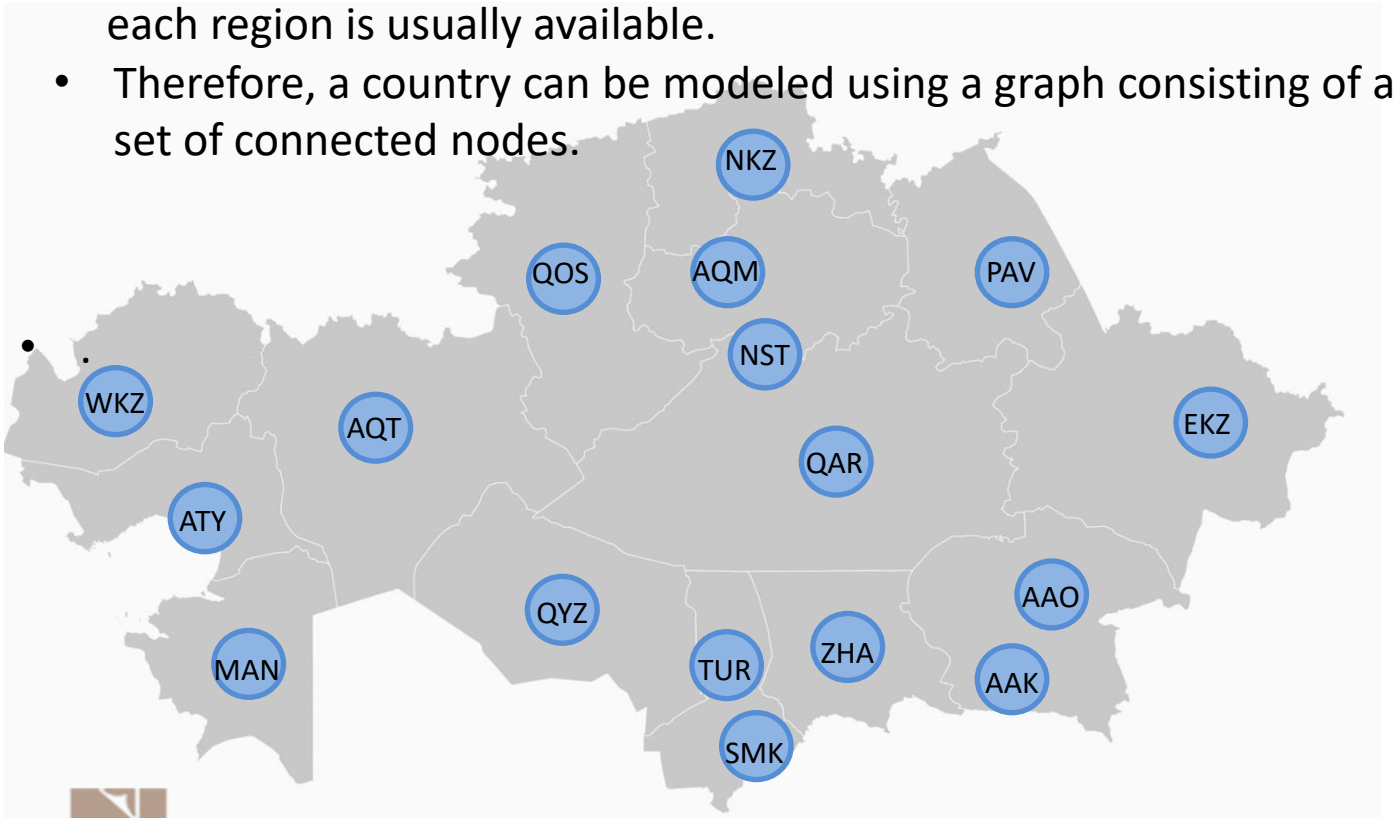
Lombardy Simulation Results



- The model fits available data well.
- The epidemic might last till June.
- The number of deaths in Lombardy can reach 15,000.
- We will use Lombardy model parameters as a baseline for Kazakhstan nodes.

Kazakhstan As a Network

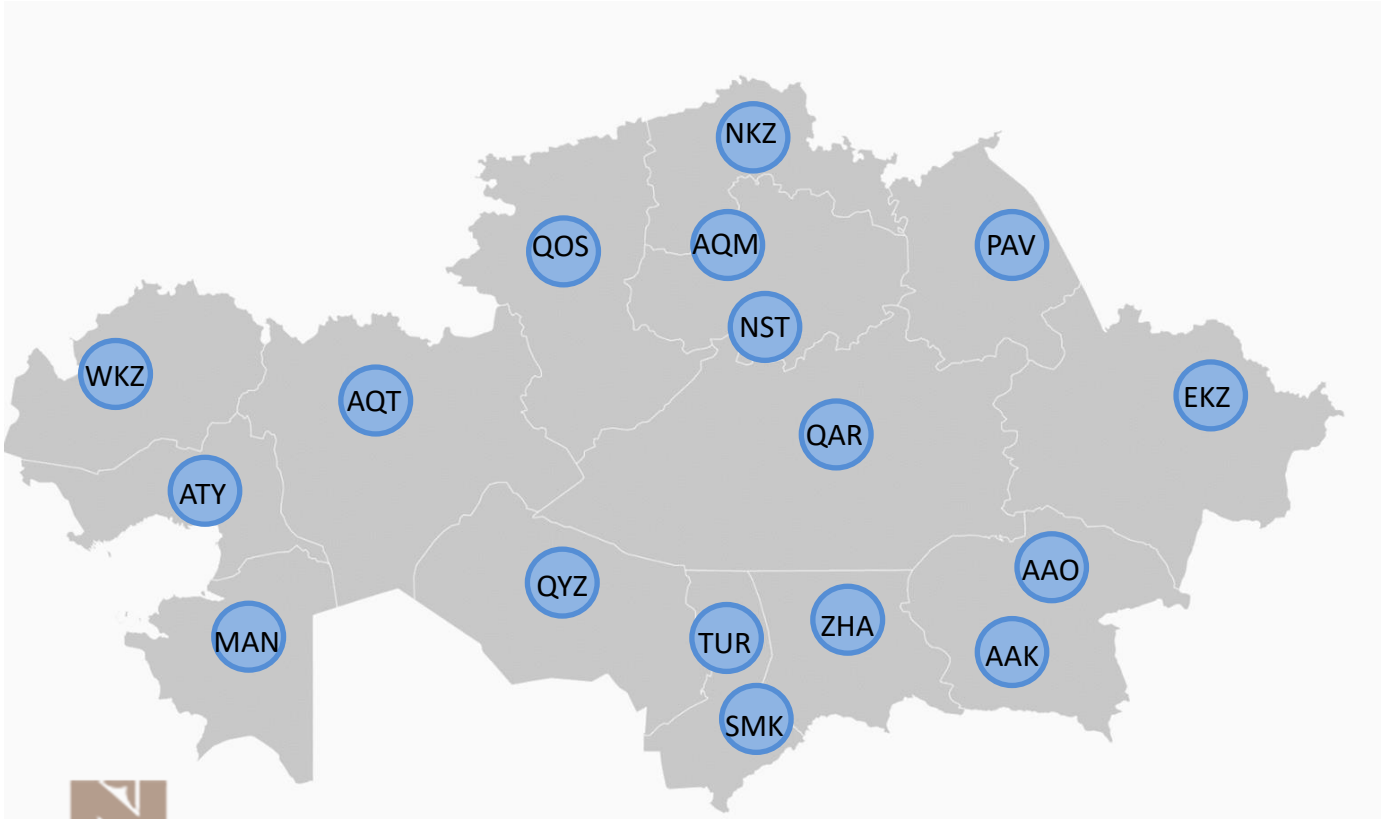
- During epidemics, measures are usually taken for specific administrative regions (for example, restricting movement within or between regions).
- These regions are also connected via air, railway and highway networks.
- Epidemics start usually in hubs with the import of the disease from another country and propagates to other regions via transportation between regions.
- Statistical data on the population, hospital capacity, and transportation connections of each region is usually available.
- Therefore, a country can be modeled using a graph consisting of a set of connected nodes.



Administrative Region	Population
Almaty	2039376
Almaty Qalasy	1854556
Aqmola	738587
Aqtobe	869603
Atyrau	633801
West Kazakhstan	652314
Jambyl	1125297
Mangystau	678224
Nur-sultan	1078362
Pavlodar	753804
Qaragandy	1378554
Qostanai	872736
Qyzylorda	794165
East Kazakhstan	1378504
Shymkent	1011511
North Kazakhstan	554519
Turkistan	1981747
Kazakhstan	**1839560**

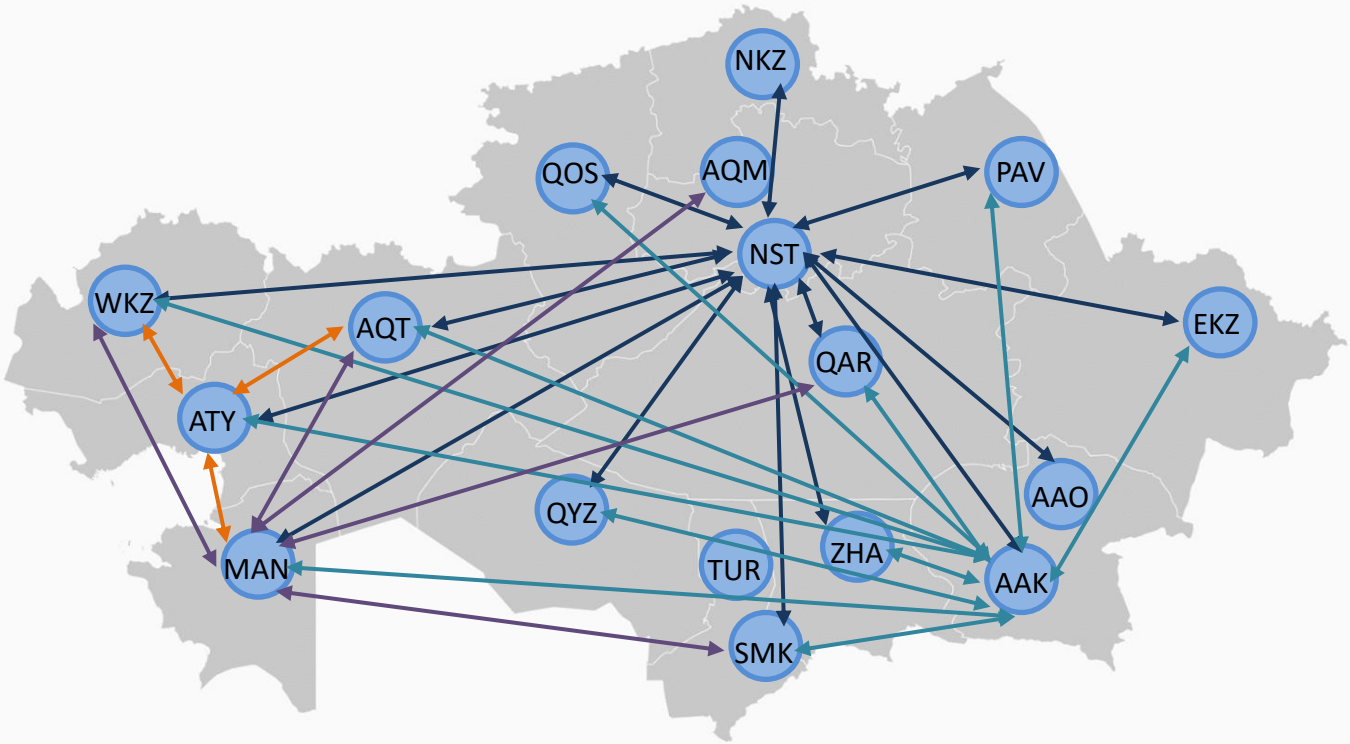
Kazakhstan As a Network

- Kazakhstan is administratively divided into 17 regions (14 oblasts and 3 cities of Republican significance: Almaty, Nur-Sultan, and Shymkent).
- We modeled Kazakhstan as a network with the regions as the nodes and the air, rail and highway connections as the edges.
- Epidemic will most likely start in a hub (Nursultan or Almaty) and start moving to other regions via transportation networks.



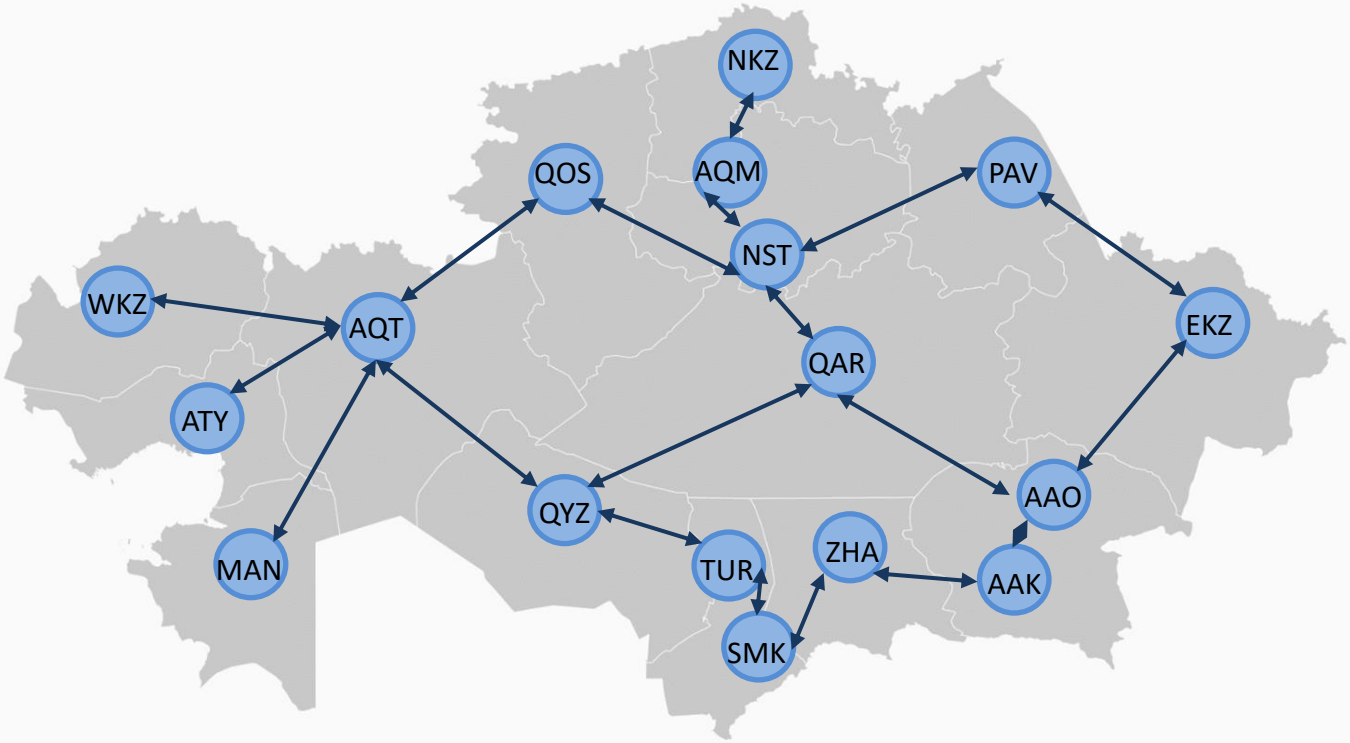
Administrative Region	Population
Almaty	2039376
Almaty Qalasy	1854556
Aqmola	738587
Aqtobe	869603
Atyrau	633801
West Kazakhstan	652314
Jambyl	1125297
Mangystau	678224
Nur-sultan	1078362
Pavlodar	753804
Qaragandy	1378554
Qostanai	872736
Qyzylorda	794165
East Kazakhstan	1378504
Shymkent	1011511
North Kazakhstan	554519
Turkistan	1981747
Kazakhstan	**18395600**

Kazakhstan Air Transport Network



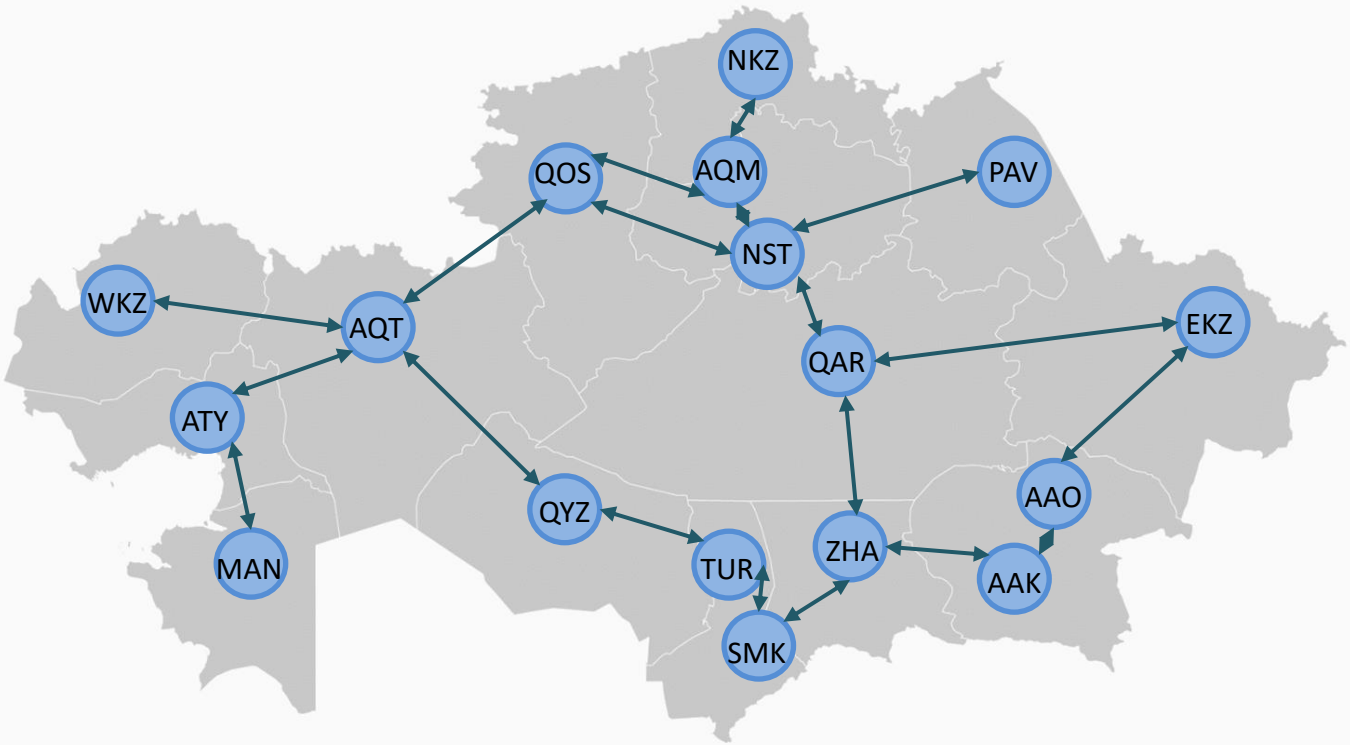
AAO	Almaty
AAK	Almaty Qalasy
AQM	Aqmola
AQT	Aqtobe
ATY	Atyrau
WKZ	West Kazakhstan
ZHA	Jambyl
MAN	Mangystau
NST	Nur-sultan
PAV	Pavlodar
QAR	Qaragandy
QOS	Qostanai
QYZ	Qyzylorda
EKZ	East Kazakhstan
SMK	Shymkent
NKZ	North Kazakhstan
TUR	Turkistan

Kazakhstan Highway Network



AAO	Almaty
AAK	Almaty Qalasy
AQM	Aqmola
AQT	Aqtobe
ATY	Atyrau
WKZ	West Kazakhstan
ZHA	Jambyl
MAN	Mangystau
NST	Nur-sultan
PAV	Pavlodar
QAR	Qaragandy
QOS	Qostanai
QYZ	Qyzylorda
EKZ	East Kazakhstan
SMK	Shymkent
NKZ	North Kazakhstan
TUR	Turkistan

Kazakhstan Railway Network



AAO	Almaty
AAK	Almaty Qalasy
AQM	Aqmola
AQT	Aqtobe
ATY	Atyrau
WKZ	West Kazakhstan
ZHA	Jambyl
MAN	Mangystau
NST	Nur-sultan
PAV	Pavlodar
QAR	Qaragandy
QOS	Qostanai
QYZ	Qyzylorda
EKZ	East Kazakhstan
SMK	Shymkent
NKZ	North Kazakhstan
TUR	Turkistan

Daily Transition Matrix Between Nodes of Kazakhstan

- We collected data from open sources on the daily air, rail and motorway travel between different regions of Kazakhstan.
- We combined this information in a daily transition matrix.

$$T_{total} = T_{air} + T_{rail} + T_{highway}$$

- In our simulation, we randomly sample the population of a node for each transfer.
- For instance, there are 1042 persons traveling from Almaty to Nursultan daily. Almaty has a population of around 2,000,000. Assuming there are 15,000 Infected person in Almaty, there would be around 8 Infected persons transferred from Almaty to Nursultan daily.

Node (from/to)	Almaty	Almaty Qalasy	Aqmola	Aqtobe	Atyrau	West Kazakhstan	Jambyl	Mangystau	Nur-sultan	Pavlodar	Qaragandy	Qostanai	Qyzylorda	East Kazakhstan	Shymkent	North Kazakhstan	Turkistan
Almaty	0	5100	0	0	0	0	0	0	1042	0	120	0	0	60	0	0	0
Almaty Qalasy	5100	0	0	824	654	668	1281	688	1194	744	1352	827	893	1532	1454	0	240
Aqmola	0	0	0	0	461	0	0	478	2057	0	180	0	0	0	0	0	0
Aqtobe	0	824	0	0	496	0	0	515	651	0	721	0	0	0	0	0	0
Atyrau	0	654	461	496	0	615	0	503	540	0	587	0	0	0	527	0	0
West Kazakhstan	0	668	0	0	615	0	0	450	550	0	0	0	0	0	0	0	0
Jambyl	0	1281	0	0	0	0	0	0	745	0	60	0	0	0	0	0	0
Mangystau	0	688	478	515	503	450	0	0	563	0	1035	0	0	0	549	0	0
Nur-sultan	1042	1194	2057	651	540	550	745	563	0	660	1770	652	798	1058	706	718	0
Pavlodar	0	744	0	0	0	0	0	0	660	0	103	0	0	0	0	0	0
Qaragandy	120	1352	180	721	587	0	60	1035	1770	103	0	766	801	1078	969	60	240
Qostanai	0	827	0	0	0	0	0	0	652	0	766	0	0	0	0	0	0
Qyzylorda	0	893	0	0	0	0	0	0	798	0	801	0	0	0	0	0	0
East Kazakhstan	60	1532	0	0	0	0	0	0	1058	0	1078	0	0	0	0	0	0
Shymkent	0	1454	0	0	526	0	0	549	706	0	969	0	0	0	0	484	900
North Kazakhstan	0	0	0	0	0	0	0	0	718	0	60	0	0	0	484	0	0
Turkistan	0	240	0	0	0	0	0	0	0	0	240	0	0	0	900	0	0

The rows and columns for hubs (Nur-Sultan and Almaty) are highlighted.

Kazakhstan COVID-19 Timeline

Day	Date	Death	Confirmed Cases	Tests	Event
-6	**3/13/2020**	**0**	**2**	**?**	Russia and Kyrgyzstan put restriction on free border crossings. Announcement of online classes. Country wide events are postponed.
-3	**3/16/2020**	**0**	**3**	**?**	Restriction on large gatherings.
0	**3/19/2020**	**0**	**44**	**?**	National emergency state. Closure of Nur-Sultan and Almaty. Restriction on working hours of restaurants. Penalties for quarantine violation are officially set.
3	**3/22/2020**	**0**	**56**	?	Lockdown of Nur-Sultan and Almaty.
7	**3/26/2020**	**1**	**81**	?	First death. Restrictions of movement in Shymkent. Limitation to go outside in Nur-Sultan and Almaty except for essentials.
11	**3/30/2020**	**1**	**293**	?	Lockdown of Aktobe, Kostanai, Atyrau, Turkistan, Karaganda regions.

1. Assuming average duration of 19 days for the disease, the first person who passed away should have contracted it on around 7 March 2020.
2. The first deaths in Lombardy were reported 33 days later than the import of COVID-19 from China.
3. Even though Almaty and Nur-Sultan are locked down since 22 March 2020, there are COVID-19 cases in other regions which cannot be traced back to confirmed cases.

- Therefore, combining (1), (2), and (3), we will assume in our simulations that there were 10 Exposed persons in both Almaty and Nursultan on 1 March 2020.

Comparing Lombardy and Kazakhstan

- **Lombardy**
  - Predicted COVID-19 arrival: 21 January 2020
  - First serious measure: 2 March 2020
  - Delay between the first case and the response is 39 days.

- **Kazakhstan**
  - Predicted COVID-19 arrival: 1 March 2020
  - First serious measure: 13 March 2020
  - Delay between the first case and the response is 12 days.

Kazakhstan was aware of the danger due to the extent of the epidemic in other countries and was quick to make interventions to prevent the spread of COVID-19.

Fortune favors the prepared.
Louis Pasteur (French biologist)

Comparing Lombardy and Kazakhstan

- **Lombardy**
  - Predicted COVID-19 arrival: 21 January 2020
  - First serious measure: 2 March 2020
  - Delay between the first case and the response is 39 days.

- **Kazakhstan**
  - Predicted COVID-19 arrival: 1 March 2020
  - First serious measure: 13 March 2020
  - Delay between the first case and the response is 12 days.

Kazakhstan was aware of the danger due to the extent of the epidemic in other countries and was quick to make interventions to prevent the spread of COVID-19.

Fortune favors the prepared.
Louis Pasteur (French biologist)