

Modelagem do Covid-19 na Cidade de Nova York

Pedro Castilho Dall

Setembro 2020

1 Introdução

1.1 Covid em Nova York

O primeiro caso de COVID-19 no estado de Nova York foi confirmado em primeiro de Março de 2020, logo em seguida o estado se tornou epicentro da pandemia global. Pouco mais de 1 mês depois do primeiro caso, Nova York possuía mais casos confirmados que qualquer outro país no mundo, com exceção dos Estados Unidos. Ao final de setembro o estado reportou um total de 10 milhões de testes com 450,473 casos positivos e 25,428 mortes. [2]

Mesmo tendo sido o estado americano mais atingido no começo da pandemia, em 17 de Junho o governador Andrew Cuomo anunciou que a taxa de infecção deles era a menor dentro do país. [2]

1.2 Linha do Tempo

Apos o primeiro caso confirmado, o Prefeito da cidade de NY tentou ignorar o vírus, encorajando os cidadãos a manterem suas rotinas usuais. Na semana seguinte, 7 de Março, o Governador anunciou estado de alerta para todo o estado. Mesmo assim, três dias depois, 10 de Março, o prefeito manteve um tom de despreocupação com o vírus afirmando que era basicamente uma gripe comum.

As divergências entre o Governador e o Prefeito, somadas à falta de preparo e à complexidade da rede de pessoas, principalmente da cidade de Nova York, fez com que o vírus se espalhasse com uma velocidade altíssima.

15 de março com 942 casos confirmados, 45% superior ao dia anterior, o Estado anuncia o fechamento das escolas. Ao final do mês o numero de casos confirmados pulou para 83,000. Neste momento o sistema de saúde começou a entrar em colapso, um dos médicos no hospital Elmhurst Hospital Center em Queens relatou a situação como "apocalíptica". [1]

Ao final de Abril o numero de casos confirmados era superior ao de 300,000 com pouco menos de 25,000 mortes. Porém neste momento a pandemia já começava a ser controlada com cerca de 1% de aumento diário de casos. [7]

Nos meses seguintes planos de reabertura foram introduzidos e tem gerado um controle bom epidemia, diferente de muitos outros estados e países, NY não foi vitima de uma segunda onda ate o momento. [3]

2 Artigos Relevantes

2.1 Dinâmica de hospitalização em Nova York

Nesse artigo, pesquisadores mostraram a eficácia do modelo SIR para estimar o número de camas ocupadas em várias cidades incluindo Nova York. A modelagem das camas ocupadas também é bem simples e é dada pela seguinte equação [5] :

$$\frac{dP(t)}{dt} = \alpha\gamma(t - \tau) - P(t)/t_H$$

Onde γ é dado por $\frac{1}{t_R} - \beta$, o primeiro termo sendo o inverso do período de recuperação e o segundo a taxa de transmissão, e t_H é o tempo médio de ocupação dos leitos.

Mesmo com um modelo muito simples os pesquisadores conseguiram encaixar parâmetros com uma alta correlação com os dados reais. um resultado interessante é que o tempo de hospitalização na cidade de NY era notavelmente menor que a maioria das cidades testadas.

2.2 Utilizando a mobilidade urbana em Nova York em modelos

O segundo artigo que exploramos critica a simplicidade dos modelos usuais de compartimentos por considerar a população homogênea e por não incluir os efeitos da mobilidade urbana. [6]

Os autores afirmam que a deficiência desses modelos poderiam ser resolvidas usando um modelo epidemiológico em grafos. Contudo outro problema surge com isso, a reconstrução de uma rede do tamanho de uma cidade é inviável, muito mais sua rede temporal. Além de outros problemas como a dificuldade de escalar essa classe de modelo.

O artigo propõe uma versão estendida do SEIR para incluir mobilidade urbana, neste modelo consideramos várias subpopulações dentro da população original. Cada uma dessas subpopulações é residente de uma zona da cidade e visitante de outra zona, cada uma das subpopulações podem ser encontradas de 4 formas, em suas próprias zonas, em atividades em outras zonas, em trânsito ou em trânsito de volta para suas zonas.

Com essas definições os compartimentos de suscetíveis, expostos, infectados e recuperados ganham muito mais estrutura. O artigo explorou esse novo modelo na cidade de NY com dados sobre a mobilidade urbana e apontou que o contágio total durante a viagem pode ser responsável por 28,8% do total de casos em NYC em 26 de março de 2020.

2.3 Levando em conta diferentes faixas etárias

O terceiro artigo apresenta outra variação do modelo SEIR, em particular o K-SEIR, onde o modelo tradicional é subdividido em K grupos. Os autores do artigo queriam investigar o efeito de quarentenas parciais dependendo da faixa etária. Eles também incluíram número de camas disponíveis em hospitais e CTIs. [4]

A ideia básica por trás é utilizar as mesmas equações do modelo SEIR porém em K grupos distintos e depois utilizar uma matriz de mistura para modelar a quantidade e taxa de interações entre diferentes faixas etárias.

Os autores afirmam que em país com um sistema de saúde desenvolvido quarentenas por faixa etária possuem uma eficácia 3 vezes maior no quesito salvar vidas e poupar recursos se comparado à quarentena total adotada em muitos países.

3 Metodologia

3.1 Modelagem

Para modelar a epidemia da COVID-19 na cidade de Nova York utilizaremos um modelo compartimentado da forma SEIR com modificações. Para tentar aproximá-lo o máximo da realidade incluiremos nele hospitalizados e infectados assintomáticos .

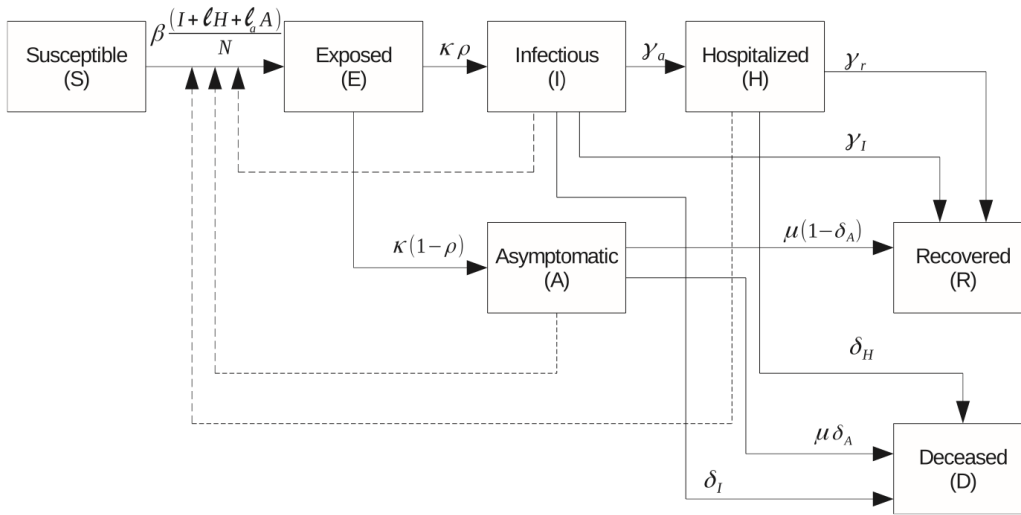
Compartimento	Descrição
S	Suscetíveis
E	Expostos
I	Infectados
A	Assintomáticos
H	Hospitalizados
R	Recuperados
D	Falecidos

Temos duas classes de parâmetros no modelo, constantes e parâmetros variáveis no tempo. As constantes costumam depender principalmente das características do patógeno e podemos assumir que, pelo menos no curto prazo, elas se mantêm constantes. A segunda classe de parâmetros são as que variam com o tempo, elas costumam de depender de medidas publicas e conhecimentos dos médicos. [5]

Parâmetros	Descrição
$\beta(t)$	taxa de transmissão de pessoa para pessoa por unidade de tempo (dia)
$l(t)$	transmissibilidade relativa de pacientes hospitalizados
$l_a(t)$	transmissibilidade relativa de infectados assintomáticos
κ	taxa na qual um indivíduo deixa a classe exposta ao se tornar infeccioso
ρ	proporção de progressão da classe E exposta para classe I infectada sintomática
γ_a	taxa na qual indivíduos sintomáticos são hospitalizados
γ_i	taxa de recuperação sem ser hospitalizado
γ_r	recovery rate of hospitalized patients
μ	taxa de infecciosos assintomáticos que não transmitem mais
δ_A	proporção de progressão da classe A assintomática para a classe D falecida
δ_H	taxa de mortalidade de pacientes hospitalizados
δ_I	taxa de mortalidade de pacientes infectados

O modelo segue o seguinte comportamento[5]:

$$\begin{aligned}
\dot{S} &= -S \frac{\beta(I + \ell_a A + \ell H)}{N}, \\
\dot{E} &= S \frac{\beta(I + \ell_a A + \ell H)}{N} - \kappa E, \\
\dot{I} &= \kappa \rho E - (\gamma_a + \gamma_I + \delta_I) I, \\
\dot{A} &= \kappa(1 - \rho) E - \mu A, \\
\dot{H} &= \gamma_a I - (\gamma_r + \delta_H) H, \\
\dot{R} &= \gamma_r I + \gamma_r H + \mu(1 - \delta_A) A, \\
\dot{D} &= \delta_H H + \delta_I I + \mu \delta_A A.
\end{aligned}$$



3.2 Análise dimensional e Adimensionalização

Utilizaremos análise dimensional para confirmar que as interpretações de cada um dos parâmetros descritos acima. Sabemos que todas os compartimentos representam populações, com isso dados a seguinte unidade para cada compartimento:

$$[S] = [E] = [I] = [R] = [Q] = [A] = [D] = \text{Pessoas}$$

Em análise dimensional podemos tratar derivadas como razões, com isso a dimensão dos diferenciais é:

$$\left[\frac{dS}{dx}\right] = [S]/[x] = \text{Pessoas} \cdot T^{-1}$$

Utilizando as equações diferenciais da modelagem descobrimos que β tem dimensão $1/(\text{Pessoas} \cdot T)$, o que faz sentido pois é uma taxa que depende de da quantidade de pessoas e do tempo que o vírus esta se espalhando.

$$\text{Pessoas} \cdot T^{-1} = [\beta][S][(I + A + r(x)(I_i + A_i))] \Rightarrow \text{Pessoas} \cdot T^{-1} = [\beta](\text{Pessoas})^2 \Rightarrow [\beta] = T^{-1} \cdot (\text{Pessoas})^{-1}$$

Utilizando as equações diferenciais dos infectados descobrimos que:

$$[\tau][E] = [\sigma][I] = [\rho][I] = Pessoas \cdot T^{-1} \Rightarrow [\tau] = [\sigma] = [\rho] = T^{-1}$$

O que também está de acordo com nossas intuições, o inverso de períodos e taxas que dependem somente do tempo devem possuir essa dimensão de fato. Os outros parâmetros são derivados de maneira análoga.

4 Resultados

Durante a elaboração deste trabalho enfrentei muitas dificuldades com o primeiro modelo (Presente no artigo feito na A1) que implementei, acabei ignorando o alerta de George Box e me apaixonei pelo modelo que estava tentando ajustar. Acabei errando diversas vezes por conta disso, mas sem dúvida alguma foi a parte mais recompensadora do trabalho, tive a oportunidade de estudar a fundo coisas que vão muito além do curso por conta disso e quero poder compartilhar meus erros com o primeiro modelo além dos resultados do segundo modelo. Muito provavelmente meus erros serão muito extensos para poder colocar de forma sucinta no trabalho por isso envio junto o jupyter notebook não só dos resultados mas também a do primeiro modelo com erros comentados.

4.1 Modelos vs Realidade e como overfittar:

”Todo modelo está errado, mas alguns são muito úteis.”

- George Box

Ao explorar modelos epidemiológicos nos deparamos a seguinte dúvida: Complexidade ou simplicidade?

Modelos mais complexos permitem uma flexibilidade maior para os resultados, permitindo que exista de fato uma representação fiel dos dados. Contudo isso vem justo ao custo computacional e até mesmo impossibilidade de achar o mínimo global.

Essa foi uma das dificuldades enfrentadas no primeiro modelo apresentado, ao introduzir um total de 11 equações diferenciais com 16 parâmetros distintos achar o mínimo global extremamente difícil e facilitamos a possibilidade de overfitting ao cair em um mínimo global que aproxima relativamente bem os dados mas com parâmetros que não fazem sentido biológico.

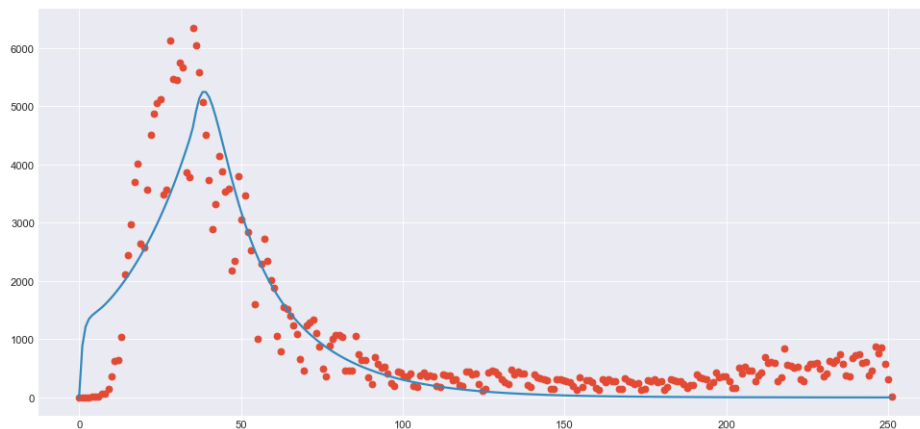
Modelos mais simples permitem que encontremos respostas com mais facilidade além de, em alguns casos, diminuir possíveis parâmetros overfittados.

Um exemplo claro de overfit:

Por mais nossos parâmetros produzam uma curva que lembre os dados, é óbvio que ela não é uma boa representação da realidade. Toda epidemia começa com um crescimento exponencial, então não faz sentido que nosso modelo também não possua essa propriedade, ao tentar interpretar o que está acontecendo percebemos o quão errado ele está, imagine um cenário onde nos primeiros dias da pandemia o número de infectados cresce mais rapidamente que qualquer outro dia, no mínimo pouco intuitivo e estranho.

A receita para uma estimação péssima dos dados é simples e você pode fazer com ingredientes encontrados em casa:

1. Tentar estimar os parâmetros somente com o número de infectados, o que é um problema ainda maior quando seu modelo não possui identificabilidade, o que provavelmente era o caso do primeiro modelo usado para entender a atual pandemia.



2. Aumentar de forma desnecessária o espaço paramétrico. Na minha busca insaciável pelo ajuste perfeito deixei de lado o bom senso junto com a navalha de Ockham (nunca faça isso) e aumentei sem nenhuma parcimônia a flexibilidade de muitos parâmetros permitindo que eles variassem no tempo, para isso tive que escolher um espaço funcional onde esse parâmetro existiria. O exemplo acima foi feito escolhendo um polinômio do 4 grau para representar o percentual de pessoas praticando distanciamento social.
3. Não escolher corretamente os valores iniciais, como nosso paramétrico é extremamente grande, a chance de cairmos em um mínimo local não depende muito do chute inicial, no caso acima escolhemos qualquer valor inicial e acabamos chegando nessa atrocidade.

4.2 Estimação de Parâmetros

Como estamos usando um modelo determinístico, ficamos muito limitados na modelagem da segunda onda. Quando modelos semelhantes a este, regidos por um sistema de EDOs, entram em seus estados estacionários é impossível que eles saiam sem uma intervenção externa nos compartimentos. Para compensar por essa falha no modelo separamos nossa pandemia em 3 intervalos:

Primeira onda, Platô e Segunda onda.

Começamos nosso modelo com 10 infectados e estimamos os parâmetros de nosso modelo durante a primeira onda. Utilizamos os valores finais dos compartimentos durante a primeira onda para construir o platô. e utilizamos os mesmos valores como condições iniciais para modelar a segunda onda

Vale ressaltar que estamos ajustando nosso modelo não somente aos dados de infecções diárias registradas mas também ao número de mortes acumuladas e ao número de hospitalizados na cidade de Nova York.

caso o ajuste fosse feito às observações dos novos infectados diários, veríamos que o modelo se encaixaria melhor para essa curva específica, contudo ao ajustar para os 3 dados evitamos encontrar parâmetros não realistas.

Nosso modelo consta com um total de 12 parâmetros, sendo que 3 deles são variáveis no tempo e δ_A fixamos como 0 pois assintomáticos não podem vir a falecer por conta da doença, dessa forma aumentamos nosso espaço de parâmetros de 11 dimensões para 15.

Para conseguir encontrar um bom ajuste ao modelo, um simples algoritmo de mínimos quadrados não irá ajudar muito. Para ajustar nosso modelo aos dados utilizamos Limited-memory BFGS, um algoritmo

de otimização quasi-newtoniano, para minimizar nossa função objetivo, no nosso caso escolhemos os erros quadráticos.

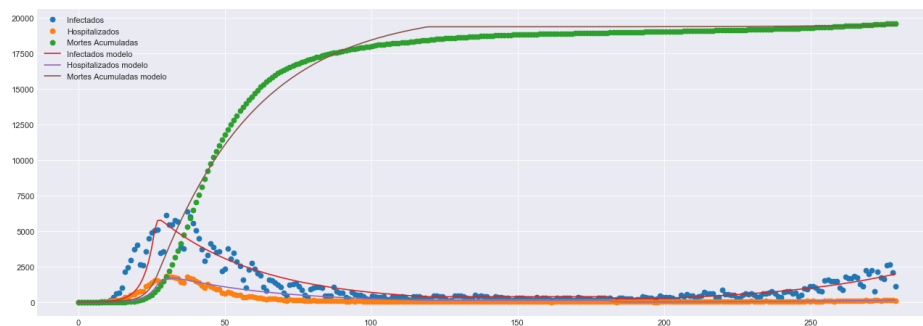
Na tabela abaixo encontramos os parâmetros estimados para a primeira onda. Os parâmetros variáveis foram representados por meio de uma função simples onde antes de τ temos que eles são constantes com o primeiro valor da tupla e após τ constantes com o segundo valor da respectiva tupla. Também estimamos o valor de τ .

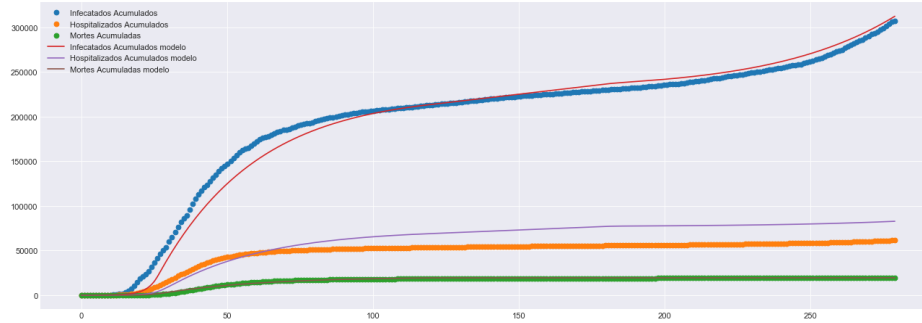
Parâmetros	Descrição
$\beta(t)$	(0.48708694, 0.19546565)
$l(t)$	(4.70218998, 0.83001964)
$l_a(t)$	(0.91713621, 0.84594214)
κ	1.17416901
ρ	0.37743780
λ_a	0.12746362
λ_i	0.02620808
λ_r	0.33274822
μ	0.23652783
δ_I	0.07108074
δ_H	0.06009496
τ	26

Abaixo temos os valores estimados para a segunda onda:

Parâmetros	Descrição
β	0.39604306
l	3.39516364
l_a	0.86270424
κ	0.88648104
ρ	0.30568232
λ_a	0.01755524
λ_i	0.71867264
λ_r	0.22374554
μ	0.25180579
δ_I	0.00163580
δ_H	1.4912e-06

Abaixo temos o modelo utilizando os parâmetros estimados:





É interessante notar como o valor de β , l e l_a variam muito dentre os 3 valores que eles assumem. No começo da pandemia eles possuíam um valor muito alto o que significa que o vírus estava se transmitindo com muita facilidade entre os indivíduos da população e após 26 dias do começo da pandemia esses valores caíram muito, mostrando que o distanciamento social e quarentena realmente tiveram um efeito muito grande para controlar a doença em Nova York.

De forma não surpreendente podemos ver que os mesmos parâmetros aumentaram, o que ajuda a explicar o surgimento da segunda onda.

4.3 Número básico de reprodução

Para cada modelo existe uma equação que descreve seu R_0 , o número básico de reprodução, um indicio de o qual infecciosa uma doença é. Existem muitas interpretações para esse parâmetro, uma forma de entender ele é como o número infecções secundárias dado que toda a população começa suscetível.[5]

Abaixo temos a formula do R_0 para o modelo apresentado:

$$R_0 = \beta \left(\frac{\rho}{\gamma_A + \gamma_I + \delta_I} + l_a \frac{1 - \rho}{\mu} + l \frac{\gamma_a \rho}{(\gamma_r + \delta_H)(\gamma_A + \gamma_I + \delta_I)} \right)$$

Substituindo os valores dos parâmetros estimados anteriormente encontramos que o R_0 em NY foi de 3.24. Ou seja, no começo da pandemia tínhamos que uma pessoa infectava em media 3.24 pessoas.

O passo a passo das contas para encontrar o R_0 pode ser encontrada no jupyter Notebook.

4.4 Análise de sensibilidade

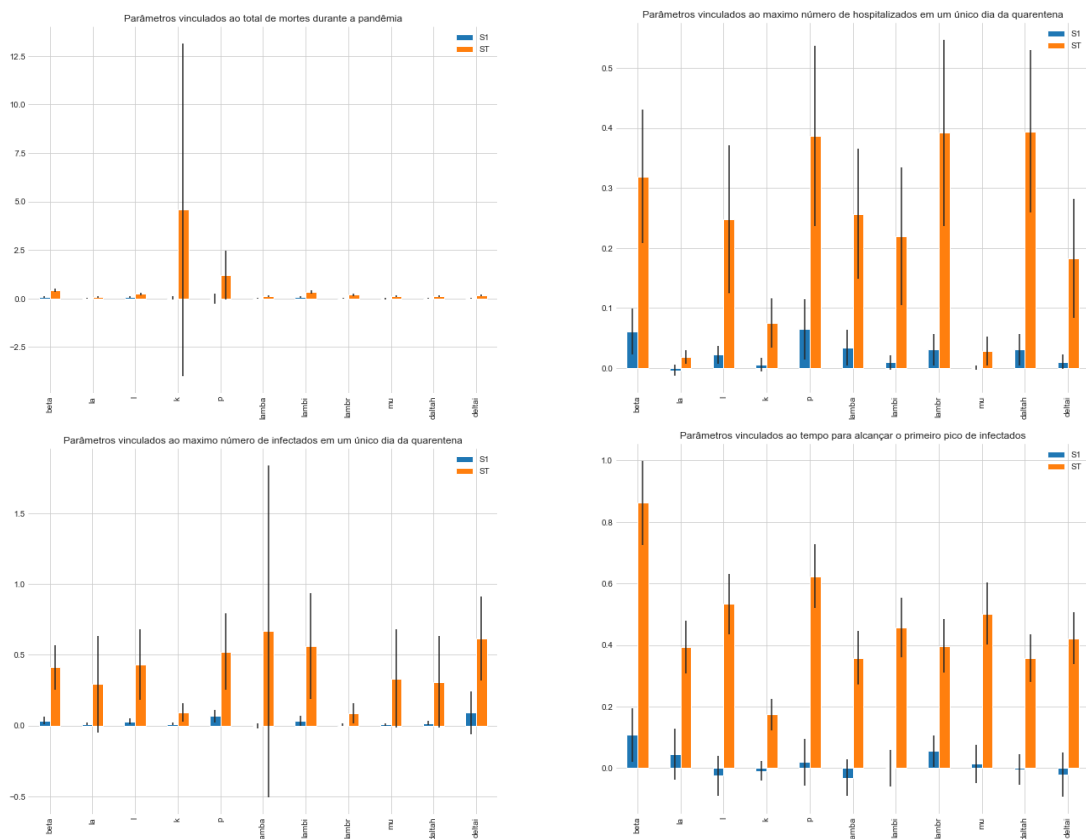
Análise de sensibilidade é o estudo de como os parâmetros de nosso modelo afetam a saída dele. Existem diversos motivos pelos quais gostaríamos de poder entender o quão importante um parâmetro é em alguma situação especificada, no caso da Covid em particular é que com uma boa noção da influência dos parâmetros as autoridades podem tomar decisões embasadas.

Quatro eventos que eu acredito serem interessantes para o conhecimento das autoridades são:

1. Número de mortes causadas pela pandemia
2. Máximo de pessoas infectadas em um único dia
3. Máximo de pessoas sendo hospitalizadas em um único dia
4. Tempo até o pico de infectadas

Muitos dos parâmetros são intrínsecos da doença e não podemos fazer nada para mudá-los, contudo outros, como a taxa de transmissão da doença, são possíveis de controlar com o medidas governamentais.

Abaixo temos os gráficos que mostram a importância de cada parâmetro para cada cenário:



Pelos nossa análise de sensibilidade parece que nossa única esperança de impedir o colapso do sistema de saúde é controlando o β , a taxa de transmissão do vírus de pessoa para pessoa. Ou seja mantendo o distanciamento social.

Percebemos também como a proporção de assintomáticos influencia no numero de hospitalizados. Por um lado se uma população tem uma quantidade muito grande de assintomáticos temos menos pessoas para serem internadas contudo o vírus se espalha mais rapidamente, é difícil dizer qual seria a pior proporção de assintomáticos sem uma análise mais detalhada.

A forma mais eficaz de diminuir o numero de óbitos ate o final da pandemia é controlando a taxa de mortes. Com o passar do tempo médicos vão entendendo melhor o vírus e sabem lidar melhor com os pacientes isso por si só já diminui a letalidade da doença.

4.5 Conclusão

Fazer uma análise completa da covid-19 não é uma tarefa fácil, ao começar com a escolha do modelo. Tantos modelos complexos quanto os simples apresentam suas desvantagens e é preciso ter muito conhecimento para dizer o quanto de cada é ideal para poder representar a realidade da melhor maneira possível.

Fizemos o possível para modelar a pandemia mas sabemos das limitações de nosso modelo. Testagens,

mobilidade urbana e grupos etários não são representados pelo modelo e isso o afasta bastante da realidade, mesmo assim conseguimos extrair conclusões úteis.

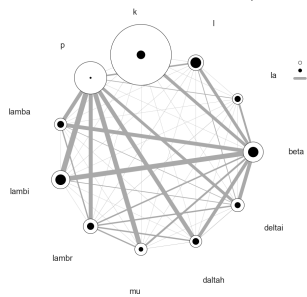
Um fato confirmado pelo modelo é o quanto o afastamento social e quarentenas afetam a taxa de transmissão do vírus, o parâmetro τ foi estimado com uma precisão altíssima, errando somente por alguns dias quando a cidade de NY começou a se preocupar com a pandemia. É óbvio que o distanciamento social diminui a taxa de transmissão do vírus, contudo hoje em dia o óbvio precisa ser repetido várias vezes e mesmo assim nem sempre é aceito.

References

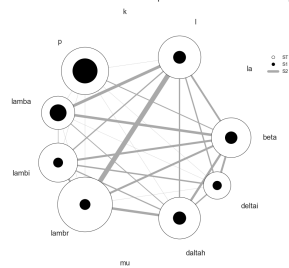
- [1] *COVID-19 Testing*. URL: <https://coronavirus.health.ny.gov/covid-19-testing>.
- [2] Yelena Dzhanova. *New York state now has more coronavirus cases than any country outside the US*. Apr. 2020. URL: <https://www.cnbc.com/2020/04/10/new-york-state-now-has-more-coronavirus-cases-than-any-country-outside-the-us.html>.
- [3] Joseph Goldstein and Jesse Mckinley. *Second Case of Coronavirus in N.Y. Sets Off Search for Others Exposed*. Mar. 2020. URL: <https://www.nytimes.com/2020/03/03/nyregion/coronavirus-new-york-state.html>.
- [4] Semu Kassa, John Boscoh Njagarah, and Yibeltal Terefe. *Analysis of the mitigation strategies for COVID-19: from mathematical modelling perspective*. Apr. 2020. DOI: 10.1101/2020.04.15.20066308.
- [5] Gregory Kozyreff. *Hospitalization dynamics during the first COVID-19 pandemic wave: SIR modelling compared to Belgium, France, Italy, Switzerland and New York City data*. 2020. arXiv: 2007.01411 [q-bio.PE].
- [6] Xinwu Qian and Satish V. Ukkusuri. *Modeling the spread of infectious disease in urban areas with travel contagion*. 2020. arXiv: 2005.04583 [physics.soc-ph].
- [7] Briana Supardi. *Albany County says there are 2 confirmed cases of coronavirus*. Mar. 2020. URL: <https://cbs6albany.com/news/coronavirus/albany-county-says-there-are-2-cases-of-coronavirus>.

5 Apêndice

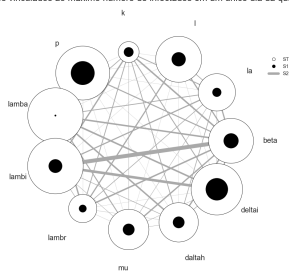
Parâmetros vinculados ao total de mortes durante a pandemia



Parâmetros vinculados ao maximo número de hospitalizados em um único dia da quarentena



Parâmetros vinculados ao maximo número de infectados em um único dia da quarentena



Parâmetros vinculados ao tempo para alcançar o primeiro pico de infectados

