



FABIO CHALUB
Universidade
Nova de Lisboa
chalub@fct.unl.pt

TESTAR, TESTAR, TESTAR

Há 12 bolas e uma balança de pratos. Uma bola é mais pesada. Qual o número mínimo de pesagem que permite identificar, com absoluta certeza, qual é a bola mais pesada? Quem diria que este problema, um clássico da matemática recreativa, pudesse inspirar sistemas mais baratos e eficientes de identificar casos positivos de COVID-19? É disso que trataremos hoje!

No início da década de 40 do século passado, o estatístico norte-americano Robert Dorfmann enfrentava o seguinte problema: como detetar rapidamente e de forma barata a presença de indivíduos portadores do vírus da sífilis num grupo grande de recrutas, pronto para iniciar o seu treinamento militar. Em época de guerra, todos os recursos são importantes. Assim, era necessária uma identificação individual, sem descurar o custo e a oferta limitada dos testes. Veja [1].

A sua proposta, que ficou conhecida como "testagem de grupo", foi a seguinte: divida os N indivíduos que devem ser testados em grupos de n pessoas. Junte amostras de sangue de todos os indivíduos de um dado grupo e teste a amostra resultante. Caso o resultado seja negativo, é porque estão todos saudáveis. Se o resultado for positivo, então cada um dos membros daquele grupo deve ser testado individualmente.

Um primeiro ponto importante é que o teste de sífilis, que deteta anticorpos, é altamente sensível. Uma pequena quantidade de antígenos é suficiente para que o resultado seja positivo. Outro ponto importante é que a probabilidade de dois recrutas estarem contaminados é considerada independente.

Assim, suponha que a probabilidade de uma dada pessoa na população estar contaminada é dada por um

número entre 0 e 1, chamado p . Portanto, a probabilidade de um único indivíduo não ter sífilis é $1 - p$, e a probabilidade de o teste de todo um grupo dar positivo é $1 - (1 - p)^n$, onde $(1 - p)^n$ é a probabilidade de que todos os indivíduos deste grupo estejam negativos. Finalmente, o número necessário de testes é dado por

$$\frac{N}{n} + n \frac{N}{n} [1 - (1 - p)^n],$$

onde a primeira parcela indica a quantidade de testes na primeira rodada (um por grupo) e a segunda expressão quantifica os testes realizados na segunda volta, onde são testados todos os indivíduos dos grupos que tiveram resultado positivo.

O número de testes necessários em relação ao tamanho do grupo, utilizando este algoritmo, é minimizado quando os grupos têm tamanho inversamente proporcional à raiz quadrada da prevalência da doença, supondo que esta seja baixa. Neste caso, o número necessário de testes por indivíduo será aproximadamente o dobro da raiz quadrada da prevalência, produzindo uma enorme economia de escala na testagem, sem perder a capacidade de identificar individualmente os casos positivos. Por exemplo, numa doença com prevalência de 1%, as testagens devem ser organizadas em grupos de

dez pessoas, com economia de 80% dos recursos necessários.

Partindo desta ideia, diversos grupos de investigação propuseram mecanismos que permitem testagens em larga escala contra a COVID-19. Numa epidemia com tantos assintomáticos e onde os testes têm sido limitados, fazê-lo de forma eficiente é fundamental. Neste contexto, aparece o artigo [2], que, com a utilização de ideias geométricas e combinatoriais, consegue produzir formas muito eficiente de testagem.

A maior parte dos autores está sediada no Ruanda, onde as ideias que apresentaremos abaixo já estão implementadas para verificação de rotinas em universidades, com apoio do Governo local, e, além disto, de acordo com os autores, foi eficiente em minimizar o impacto económico da pandemia no turismo.

Vamos começar com uma situação simples.

Imaginemos um conjunto de N pessoas, onde há apenas uma infetada. Considere todos os indivíduos organizados numa malha quadrada, organize um grupo de testagem em cada linha e em cada coluna. Apenas um grupo terá resultado positivo em cada dimensão de testagem, seja na horizontal ou na vertical. O indivíduo infetado estará na interseção, única, entre estes dois grupos. Serão necessários aproximadamente $2\sqrt{N}$ testes. Veja a figura 1.

As ideias anteriores podem facilmente ser generali-

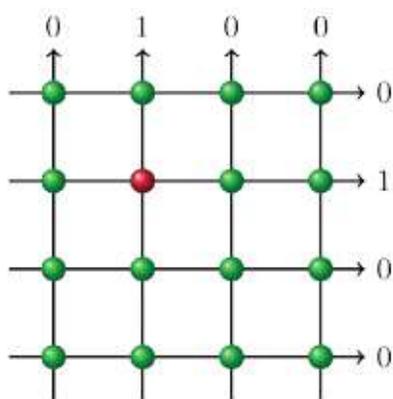


Figura 1. Veja o caso mais simples, em que os indivíduos são colocados em num arranjo quadrado bidimensional; com oito testes para 16 indivíduos (cujos resultados estão colocados ao lado das linha), é possível identificar inequivocamente quem é o único infetado.

zadas para D dimensões. A figura 2 mostra o caso $D = 3$, $N = 64$. Considere o hipercubo de lado L em que cada indivíduo é colocado nos pontos com coordenadas inteiras. Assim, temos $L^D = N$, representando os N pontos com cada coordenada dada por um número natural entre 1 e L . Depois, para cada direção e para cada valor entre 1 e L da coordenada indicativa desta direção são adicionadas as amostras dos indivíduos correspondentes e testadas. Estes são os indivíduos que estão em cada hiperplano perpendicular aos eixos cartesianos.

Desta forma, são feitos $LD = DN^{1/D}$ testes, que serão capazes de identificar perfeitamente o único indivíduo positivo entre os N do grupo inicial. Pensando em D como uma variável contínua, a expressão anterior é minimizada em $D = \log N$. Para um valor grande de indivíduos, a quantidade de testes necessária é aproximadamente dada por $e \log N$, onde $\approx 2,718$ é a constante de Euler.

A situação torna-se um pouco mais complicada quando temos dois indivíduos infetados. Vamos considerar uma direção cartesiana, e consideramos os L hiperplanos perpendiculares a esta direção. Para cada um destes, indicamos com o valor 1 caso o teste deste grupo de indivíduos dê positivo e 0 caso contrário. Assim, cada eixo cartesiano i estará associado a um arranjo de L números, 0 ou 1. Chamamos de σ_i à soma destes valores. Assim, para todo o $i \in \{1, 2, \dots, D\}$, temos $0 \leq \sigma_i \leq L$.

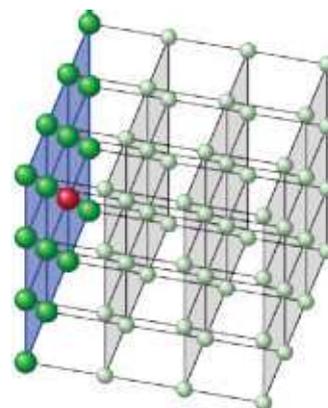


Figura 2. Há um único infetado, neste arranjo tridimensional de $4^3 = 64$ indivíduos. Ao considerar os testes dos grupos arranjados em hiperplanos paralelos, encontramos na direção marcada no desenho um primeiro resultado positivo (no plano azul), e três negativos (cinza). Desta forma, a partir dos resultados nas três direções, conseguimos identificar o infetado. Os indivíduos do hiperplano de interesse estão em destaque.

Seja d_σ a quantidade de direções i tal que $\sigma_i = \sigma$. Evidentemente, $\sum_{\sigma=0}^L d_\sigma = D$.

Se houver apenas um infetado entre os N indivíduos do grupo, então, $d_1 = D$ e $d_\sigma = 0$ para $\sigma > 1$. Suponha que há, no grupo, dois infetados. Se for o caso de que $d_2 = 1$, então $d_1 = D - 1$ e portanto estes dois infetados estão na interseção de $D - 1$ hiperplanos, ou seja, sobre uma reta. Olhando os hiperplanos perpendiculares a esta reta, e onde $\sigma_i = 2$, concluímos as posições exatas dos casos positivos. Veja a figura 3.

Se, no entanto, $d_2 > 1$, só o que conseguimos inferir é que os dois infetados estão num mesmo hiperplano de menor dimensão. Fazendo uma nova rodada de testes, similar à que já foi feita, mas considerando as amostras apenas dos indivíduos deste hiperplano, conseguimos localizar precisamente estes casos. Veja a figura 4.

Para três ou mais infetados, a situação fica mais complicada, mas é sempre mais eficiente fazer desta forma do que testar um a um. Além disto, numa amostra de indivíduos não relacionados, com prevalência não muito grande, em testes de rotina (e não por recomendação médica), então a probabilidade de encontrar diversos casos positivos é diminuta.

Há algumas hipóteses centrais nesta análise matemática: a prevalência deve ser baixa na população em geral e não deve haver razões para que naquele grupo em particular seja diferente. Noutras palavras, não se

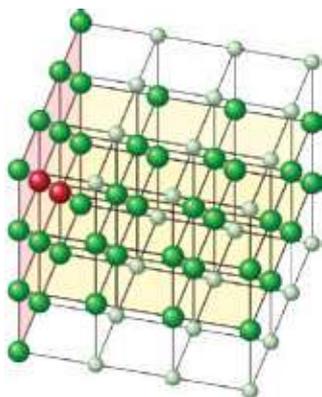


Figura 3. Temos dois infetados, no mesmo plano (rosa), mas nas outras direções estão sempre em planos distintos (amarelos, para uma das direções; a outra não está indicada). Assim $d_2 = 1$, $d_1 = 2$ e não é difícil concluir os indivíduos positivos, entre 64, com apenas 12 testes.

deve submeter a este tipo de análise grupos fortemente correlacionados, como coabitantes. Da mesma forma, não é recomendada a utilização destes algoritmos em casos de surtos, ou em testagem de pessoas sintomáticas. Mas pode ser adequado para utilização em escolas, universidades, equipas desportivas, ou até mesmo como forma de incentivo ao turismo em passageiros de avião ou navio.

Um cuidado extra que tem de ser tomado é que a diluição (ou seja, o tamanho das amostras que são levadas a laboratório) não seja tal que a análise produza falsos negativos para o grupo.

Há muitas maneiras de os matemáticos ajudarem a combater uma epidemia!

REFERÊNCIAS

[1] Dorfman, Robert. "The Detection of Defective Members of Large Populations". *Ann. Math. Statist.* 14 (1943), no. 4, 436--440. doi:10.1214/aoms/1177731363. <https://projecteuclid.org/euclid.aoms/1177731363>.

[2] Mutesa, L., Ndishimye, P., Butera, Y. et al. "A pooled testing strategy for identifying SARS-CoV-2 at low prevalence". *Nature* 589, 276–280 (2021). <https://doi.org/10.1038/s41586-020-2885-5>.

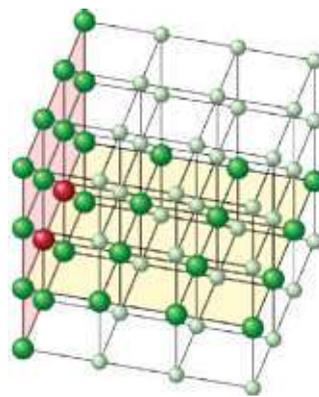


Figura 4. Nesta situação, temos dois indivíduos infectados na mesma linha paralela a um dos eixos, portanto $d_2 = 2$ e $d_1 = 1$. Para os identificar, é necessário fazer $12+8=20$ testes.