Invasion of cooperative parasites in moderately structured host populations

Cornelia Pokalyuk

joint work in progress with Vianney Brouard

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Emmy Noether-Programm
DFG Deutsche Forschungsgemeinschaft

Phages

- Phages are viruses infecting bacteria.
- Bacteria have evolved different resistance mechanisms against phages.

Image: © Dr Graham Beards All Rights Reserved

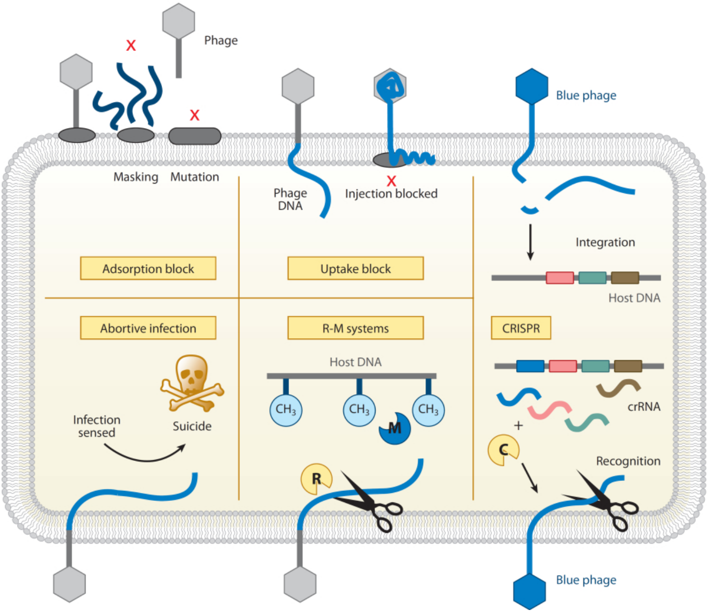
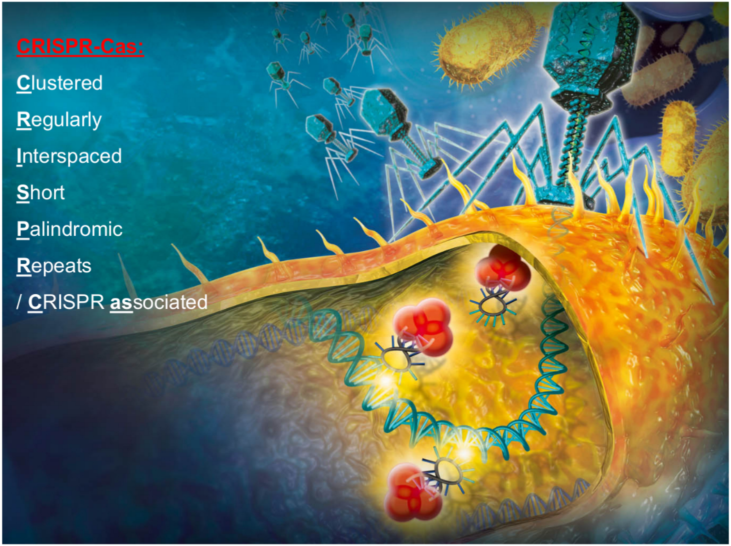


Figure: Westra et al. (2012) Annu Rev Genet.



CRISPR-Cas:

Clustered

Regularly

Interspaced

Short

Palindromic

Repeats

/ **C**RISPR **as**sociated

CRISPR and anti-CRISPR

- Bacteria evolved the CRISPR immune system to defend against phages.
- Phages evolved anti-CRISPRs (ACRs) to protect against CRISPR.

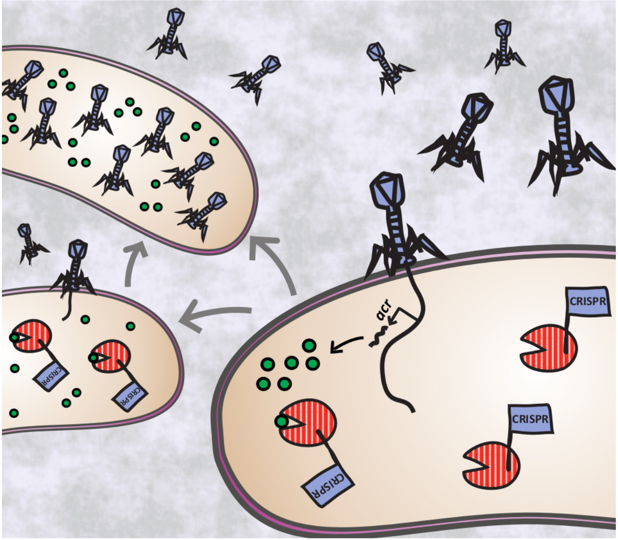


Figure by S. van Houte

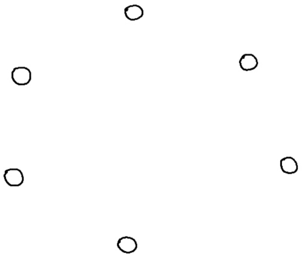
Effect of population structure

- ACR-mechanism relies on cooperation of ACR-phages
- When the density of ACR-phages is sufficiently high, simultaneous infection or reinfection of bacteria by ACR-phages is likely.
- In a well mixed-population at low frequencies ACR-phages cannot profit from cooperation.
- In a structured population co-infection of bacteria by phages is more likely.
- ACR-phage could more likely survive in structured bacterial populations.

When ACR-phages have a chance to invade a CRISPR-resistant bacterial population?

A toy model

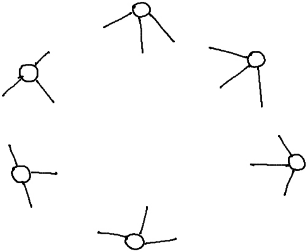
- Infections dynamics occurs on a random graph with N vertices
- Each vertex has d_N half-edges
- Half-edges are matched according to the configuration model
- The graph is fixed over the whole epidemic



- Initially on each vertex a host (CRISPR-resistant bacterium) is placed
- The infection process starts on a single host, v_N offspring parasites (Acr-phages) are produced and the host dies.

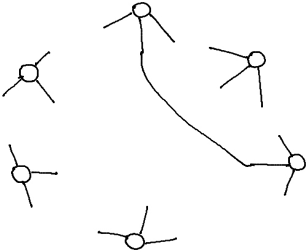
A toy model

- Infections dynamics occurs on a graph with N vertices
- Each vertex has d_N half-edges
- Half-edges are matched according to the configuration model
- The graph is fixed over the whole epidemic



- Initially on each vertex a host (CRISPR-resistant bacterium) is placed
- The infection process starts on a single host, $v = v_N$ offspring parasites (Acr-phages) are produced and the host dies.

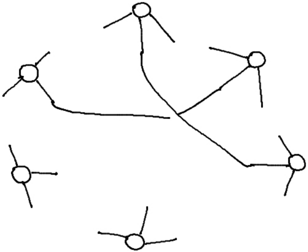
A toy model



- Infections dynamics occurs on a graph with N vertices
- Each vertex has d_N half-edges
- Half-edges are matched according to the configuration model
- The graph is fixed over the whole epidemic

- Initially on each vertex a host (CRISPR-resistant bacterium) is placed
- The infection process starts on a single host, v_N offspring parasites (Acr-phages) are produced and the host dies.

A toy model

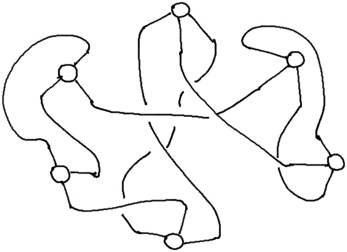


- Infections dynamics occurs on a graph with N vertices
- Each vertex has d_N half-edges
- Half-edges are matched according to the configuration model
- The graph is fixed over the whole epidemic

- Initially on each vertex a host (CRISPR-resistant bacterium) is placed
- The infection process starts on a single host, v_N offspring parasites (Acr-phages) are produced and the host dies.

A toy model

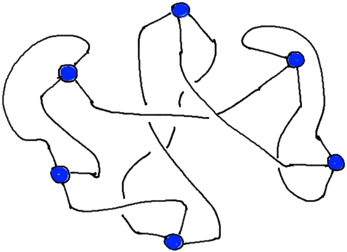
- Infections dynamics occurs on a graph with N vertices
- Each vertex has d_N half-edges
- Half-edges are matched according to the configuration model
- The graph is fixed over the whole epidemic



- Initially on each vertex a host (CRISPR-resistant bacterium) is placed
- The infection process starts on a single host, v_N offspring parasites (Acr-phages) are produced and the host dies.

A toy model

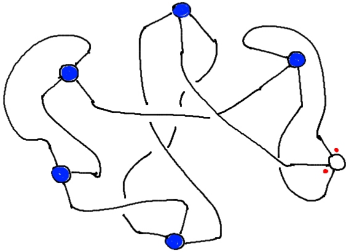
- Infections dynamics occurs on a graph with N vertices
- Each vertex has d_N half-edges
- Half-edges are matched according to the configuration model
- The graph is fixed over the whole epidemic



- Initially on each vertex a host (CRISPR-resistant bacterium) is placed
- The infection process starts on a single host, $v = v_N$ offspring parasites (Acr-phages) are produced and the host dies.

A toy model

- Infections dynamics occurs on a graph with N vertices
- Each vertex has d_N half-edges
- Half-edges are matched according to the configuration model
- The graph is fixed over the whole epidemic

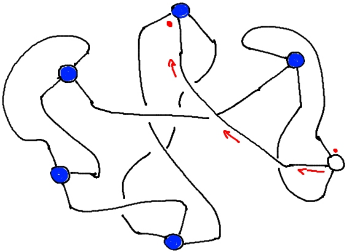


- Initially on each vertex a host (CRISPR-resistant bacterium) is placed
- The infection process starts on a single host, v_N offspring parasites (Acr-phages) are produced and the host dies.

A toy model

Reproduction

- Discrete generations: At the beginning of each generation parasites move to neighbouring vertices
- When a parasite infects a host alone, the parasite reproduces with a *small* probability p_N.
  In this case v_N offspring parasites are produced, the reproducing parasite and the host die.
  Otherwise the parasite dies and the host survives.

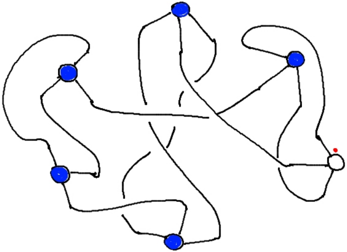


- When at least two parasites infect a host simultaneously, v_N offspring parasites are produced and the reproducing parasite and the host dies
- If a parasite moves to an empty vertex, it stays there and moves further in the next generation.

A toy model

Reproduction

- Discrete generations: At the beginning of each generation phages move to neighbouring vertices
- When a parasite infects a host alone, the parasite reproduces with a *small* probability p_N.
  In this case v_N offspring parasites are produced, the reproducing parasite and the host die.
  Otherwise the parasite dies and the host survives.

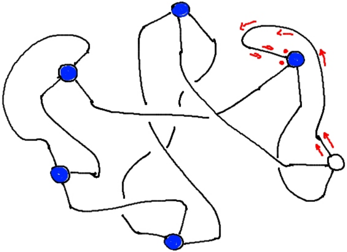


- When at least two parasites infect a host simultaneously, v_N offspring parasites are produced and the reproducing parasite and the host dies
- If a parasite moves to an empty vertex, it stays there and moves further in the next generation.

A toy model

Reproduction

- Discrete generations: At the beginning of each generation phages move to neighbouring vertices
- When a parasite infects a host alone, the parasite reproduces with a *small* probability p_N.
  In this case v_N offspring parasites are produced, the reproducing parasite and the host die.
  Otherwise the parasite dies and the host survives.

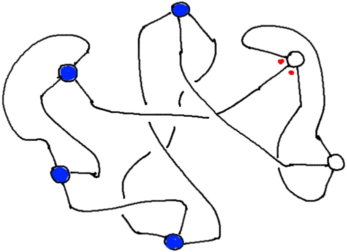


- When at least two parasites infect a host simultaneously, v_N offspring parasites are produced and the reproducing parasite and the host dies
- If a parasite moves to an empty vertex, it stays there and moves further in the next generation.

A toy model

Reproduction

- Discrete generations: At the beginning of each generation phages move to neighbouring vertices
- When a parasite infects a host alone, the parasite reproduces with a *small* probability p_N.
  In this case v_N offspring parasites are produced, the reproducing parasite and the host die.
  Otherwise the parasite dies and the host survives.

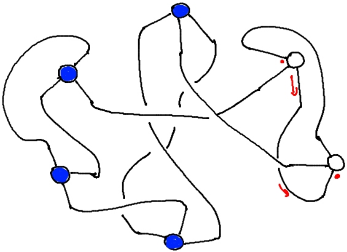


- When at least two parasites infect a host simultaneously, v_N offspring parasites are produced and the reproducing parasite and the host dies
- If a parasite moves to an empty vertex, it stays there and moves further in the next generation.

A toy model

Reproduction

- Discrete generations: At the beginning of each generation phages move to neighbouring vertices
- When a parasite infects a host alone, the parasite reproduces with a *small* probability $p = p_N$.
  In this case v_N offspring parasites are produced, the reproducing parasite and the host die.
  Otherwise the parasite dies and the host survives.



- When at least two parasites infect a host simultaneously, v_N offspring parasites are produced and the reproducing parasite and the host dies
- If a parasite moves to an empty vertex, it stays there and moves further in the next generation.

Can the parasite invade the host population?

Fix sequences $(v_N), (d_N), (p_N)$ and denote by $X_n^{(N)}$ the number of hosts killed till generation n.
Let $u > 0$ and denote by

$$E_N^u = \cup_n \left\{ \frac{X_n^{(N)}}{N} \geq u \right\}$$

the event that the parasite invades the host population to a proportion u.

Parameter regime of interest

- $p_N v_N \ll 1$:
  Reproduction by parasites hitting a host alone is subcritical
- $d_N \ll N$:
  No (almost) complete graph
- $v_N, d_N \gg 1$:
  -A large number of parasites is produced at reproduction,
  -Many hosts are reachable from a host

Theorem A

Assume

$$\text{(C1) } d_N \sim N^{\beta} \text{ for some } 0 < \beta < 1$$

$$\text{(C2) } p_N = O\left(\frac{1}{v_N^b}\right) \text{ with } b > 1$$

$$\text{(C3a) } v_N \sim c\sqrt{d_N} \text{ for some } c > 0,$$

Denote by p_{pois} the survival probability of a $Pois(\frac{c^2}{2})$-GW-process.

Case: $\frac{c^2}{2} > 1$. Let $0 < u < p_{pois}$. Then

$$\lim_{N\to\infty} \mathbb{P}(E_N^u) = p_{pois}.$$

Case: $\frac{c^2}{2} \leq 1$. Let $u > 0$. Then

$$\lim_{N\to\infty} \mathbb{P}(E_N^u) = 0.$$

Theorem B

Assume (C1), (C2) and

$$\text{(C3b) } v_N \ll \sqrt{d_N},$$

Let $u > 0$. Then

$$\lim_{N\to\infty} \mathbb{P}(E_N^u) = 0.$$

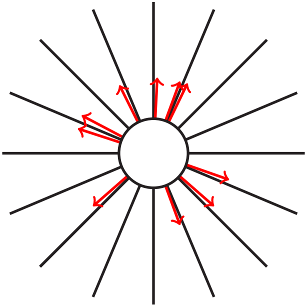
Assume (C1), (C2) and

$$\text{(C3c) } v_N \gg \sqrt{d_N}.$$

Let $0 < u < 1$. Then

$$\lim_{N\to\infty} \mathbb{P}(E_N^u) = 1.$$

Reproduction by cooperation and the birthday problem



d_N edges, v_N parasites

Reproduction by cooperation and the birthday problem

CoSame: **Co**operation from the **same** edge

A pair of parasites chooses the same edge with probability: $\frac{1}{d_N}$
Number of pairs of parasites on a vertex: $\binom{v_N}{2} \in \Theta(v_N^2)$
Number of edges occupied by a pair of parasite $\Theta\left(\frac{v_N^2}{d_N}\right)$

If $v_N \sim c\sqrt{d_N}$, the probability that two parasites attack the same vertex *by choosing the same edge* is non-trivial.

Number of pairs at the critical scaling

Consider a vertex with v_N parasites. Let q_k be the probability, that k pairs of parasites choose the same edge and the remaining parasites choose different edges.

$$q_k = \frac{\binom{v_N}{2} \cdot \dots \cdot \binom{v_N - 2(k-1)}{2}}{k!} \frac{d_N(d_N - 1) \cdot \dots \cdot (d_N - (v_N - k) + 1)}{{d_N}^k {d_N}^{v_N - k}}$$

$$\sim \left(\frac{v_N^2}{2d_N}\right)^k \frac{1}{k!} \exp\left(-\frac{\sum_{i=1}^{v_N - k + 1} i}{d_N}\right) \sim \left(\frac{v_N^2}{2d_N}\right)^k \frac{1}{k!} \exp\left(-\frac{v_N^2}{2d_N}\right)$$

If $v_N \sim c\sqrt{d_N}$,

$$q_k \sim \left(\frac{c^2}{2}\right)^k \frac{1}{k!} e^{-\frac{c^2}{2}}$$

the number of pairs of phages choosing the same edge is approximately $Pois(\frac{c^2}{2})$-distributed for large N.

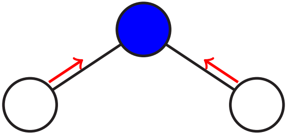
Sketch of the proof of Theorem A

Upper bound on the invasion probability

Reproduction of parasites essentially is driven only
by reproduction of parasite pairs due to CoSame
as long as the number of killed hosts is relatively small:

- When the number of new parasites is $\ell_N v_N$ about $\frac{c^2}{2}\ell_N$ new hosts get successfully infected
- ReproductionAlone: Only about $\ell_N v_N p_N \ll \ell_N$ new hosts get killed by parasites hitting a vertex alone.
- Cooperation from diffrent edges: Reproduction by parasites infecting a vertex from different edges is rare as long as the number ℓ_N of infected hosts is sufficiently small

CoDiff: Cooperation from different edges



A pair of parasites located on different edges chooses the same vertex with probability: $\frac{1}{N}$.
For $\ell_N v_N$ parasites we have roughly $\ell_N v_N$ different occupied edges and so $\Theta(\ell_N^2 v_N^2)$ pairs of occupied edges.

$\Rightarrow \Theta\left(\frac{\ell_N^2 v_N^2}{N}\right)$ new infected hosts due to CoDiff.

As long as $(\ell_N v_N)^2 \sim \ell_N^2 c^2 d_N = \ell_N^2 c^2 N^\beta \ll N$, CoDiff is rare.

Coupling with GW-process

- Numbers of killed hosts $(X_n^{(N)})$ can be estimated from above
  whp by the total size of an GW-process
  with an offspring distribution that is close to $Pois(\frac{c^2}{2})$
  until the total size of the GW population
  remains constant or crosses a level ℓ_N
  for an appropriate sequence (ℓ_N) with $(\ell_N) \to \infty$,
  such that CoDiff does not play a role.
- The probability to hit the level ℓ_N is $p_{\text{pois}} + o(1)$.
- Estimate invasion probability by 1, once the level ℓ_N is hit.

Lower bound on the invasion probability

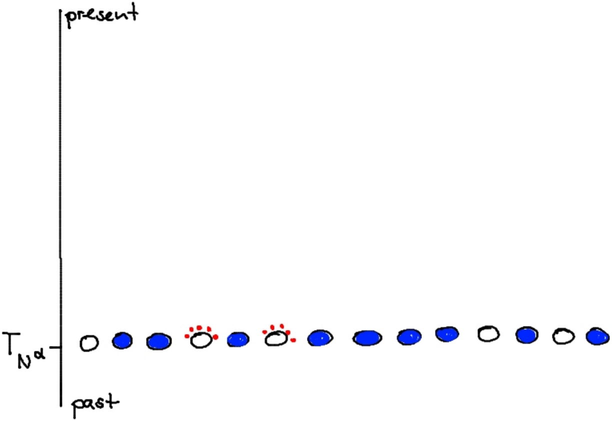
Coupling with GW-process

- For an approximation from below
  ignore RepAlone and (initially) CoDiff
- Control how likely a pair of parasites *is lost by collision*,
  that is by hitting a vertex which host has been killed before
  or is currently infected successfully by other parasites.
- Whp $(X_n^{(N)})$ can be approximated from below
  by the total population size of a GW process
  which offspring distribution is close to a Pois($\frac{c^2}{2}$)
  until the total size of the GW-process remains constant or
  crosses the level N^α for any $0 < \alpha < 1$.
- The level N^α is reached with probability $p_{pois} + o(1)$.
- Remains to show that the level uN is reached whp as well.

Growing further from the level N^{α}

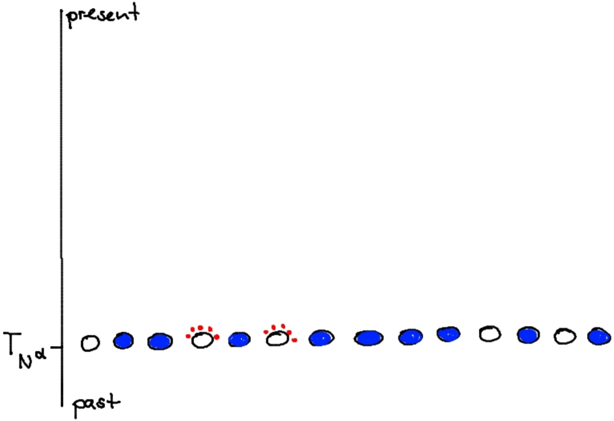
- Either ignore further CoDiff and infer potential infection routes
- Or incorporate CoDiff

Hitting the forward epidemics



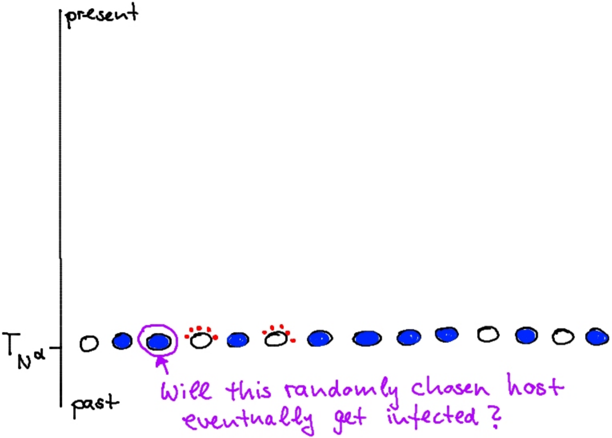
Denote by T_{N^α} the generation when the total number of killed hosts crosses the level N^α for the first time, assume $T_{N^\alpha} < \infty$.

Hitting the forward epidemic



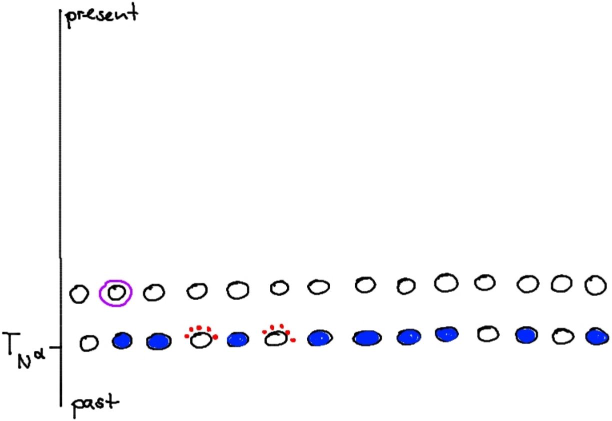
At generation T_{N^α} the number of hosts just killed is $\Theta(N^\alpha)$ (potentially more when CoDiff is not ignored).

Hitting the forward epidemic



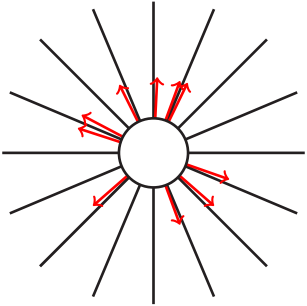
With high probability a randomly chosen individual is not infected yet at generation T_{N^α}.

Hitting the forward epidemic



Identify individuals that might infect the randomly chosen individual in the next generation

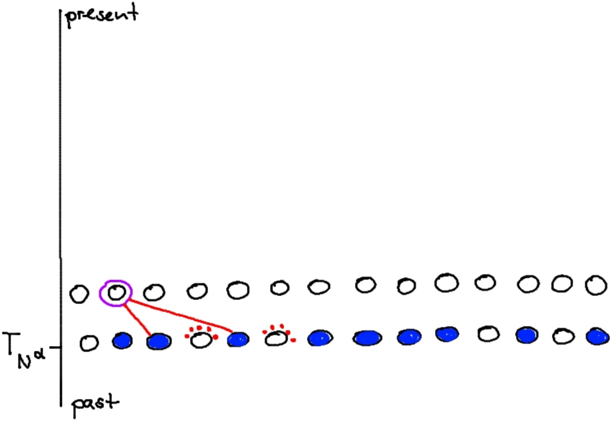
Identifying potential infection routes



Fix the parasite configurations at infection of the host over the whole epidemic.

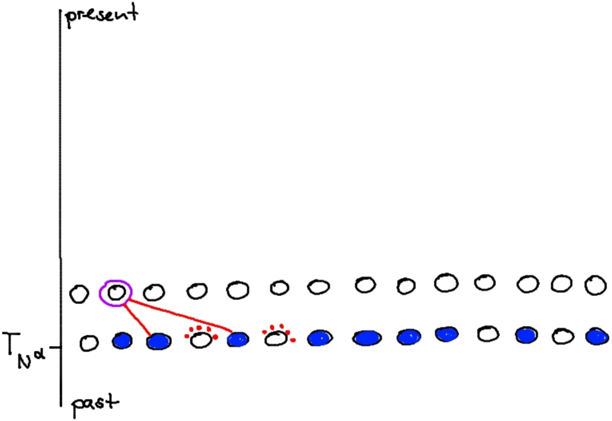
As the configuration model is also fixed over the whole epidemic, an individual can infect only a certain fixed set of other individuals.

Hitting the forward epidemic



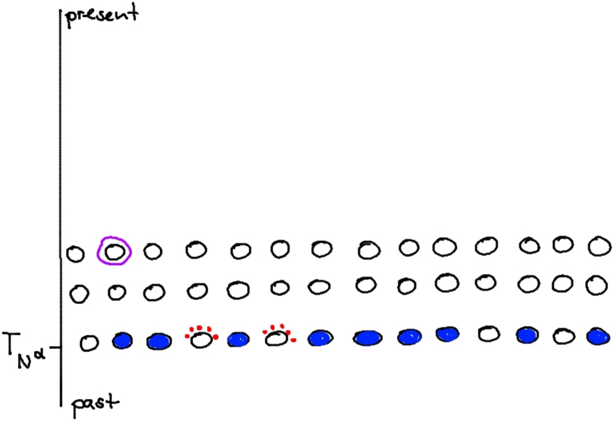
Identify individuals that might infect the randomly chosen individual in the next generation.

Hitting the forward epidemic



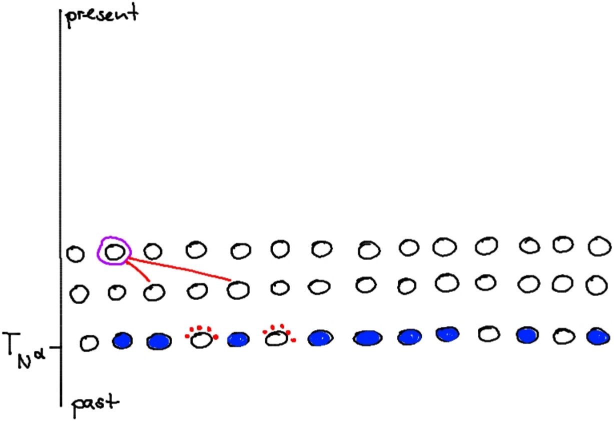
Ignoring RepAlone and CoDiff approximately Pois($\frac{c^2}{2}$)-many individuals can infect the randomly chosen individual.

Hitting the forward epidemic



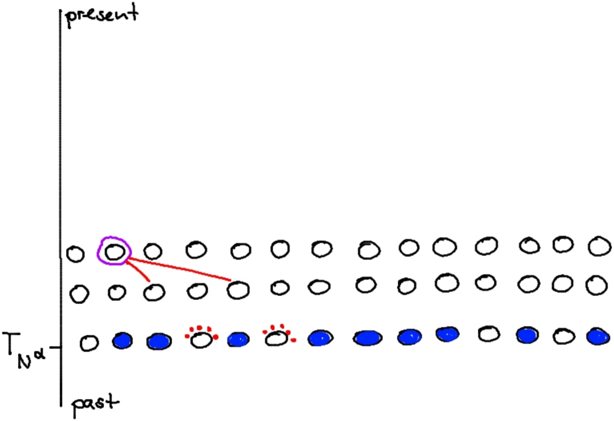
Follow the routes by which the randomly chosen individual can get infected further.

Hitting the forward epidemic



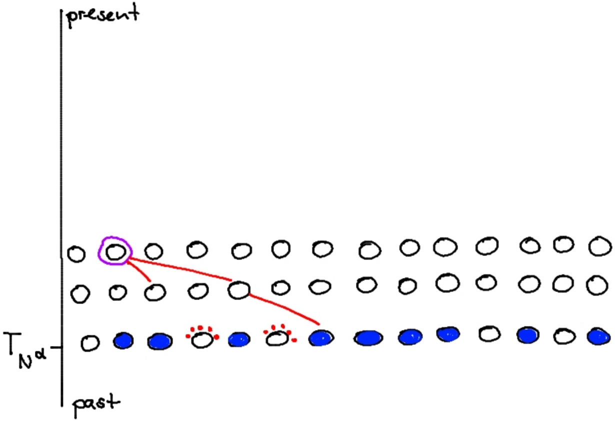
The configuration model is fixed over the total epidemic:
Infection routes can be shifted upwards.

Hitting the forward epidemic



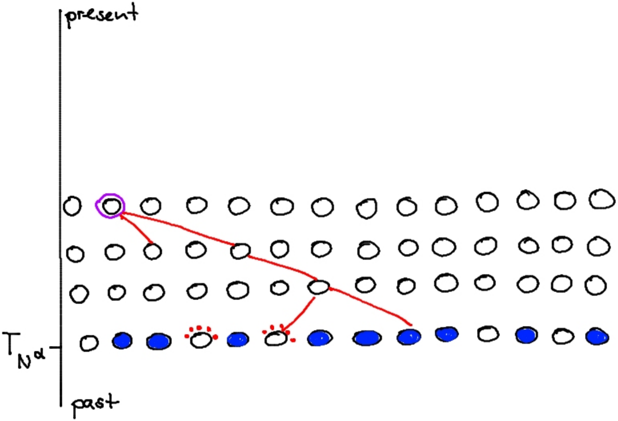
Follow infection routes until extinction
or until an infectious individual in generation T_{N^α} is hit.

Hitting the forward epidemic



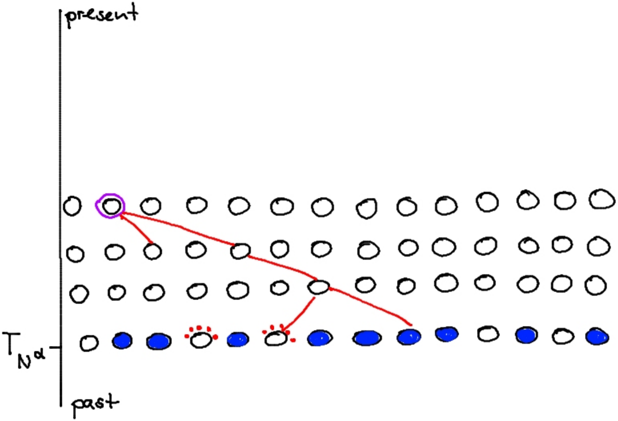
Follow infection routes until extinction
or until an infectious individual in generation T_{N^α} is hit.

Hitting the forward epidemic



Follow infection routes until extinction
or until an infectious individual in generation T_{N^α} is hit.

Hitting the forward epidemic



Throughout the backward process can be approximated from below whp by a GW-process with an approximately Pois($\frac{c^2}{2}$)-offspring distribution.

Lower bound on the invasion probability

- Consider k_N many randomly chosen hosts for a sequence $k_N \xrightarrow{N \to \infty} \infty$ sufficiently slowly
- With probability $p_{pois} + o(1)$ each chosen host gets (nearly independently) infected with probability $p_{pois} + o(1)$.
- Otherwise the number of chosen hosts that get infected is $o(k_N)$.
- Consequently, as $N \to \infty$ the proportion of eventually infected hosts is at least u for any $u < p_{\text{pois}}$ whp.

Sketch of the proof of Theorem B

- $v_N \ll \sqrt{d_N}$: It is unlikely that two or more parasites choose the same edge.
  If in addition $v_N p_N \ll 1$, the epidemic is subcritical.
  Hence, the parasite manages to invade
  the bacterial population with probability $o(1)$.
- $v_N \gg \sqrt{d_N}$: In the first generation a large number of parasites cooperate by CoSame whp.
  With high probability at least one of these parasites invades
  the bacterial population up to proportion u.

Outlook and perspective

- CoDiff allows for many additional infections after initial stochastic phase:
  Final proportion of infected bacteria is $1 - o(1)$
  with probability p_{pois}.
- Mobile bacteria: Changing configuration during the epidemic
- Not all bacteria are CRISPR-resistant,
  not all phages are Anti-CRISPR-phages
- Reproducing bacteria
- More general offspring distribution
  depending on the number of parasites
  infecting a host simultaneously
- Continuous time