M2 Project: Mathematical modeling of the cell cycle

Jan Rombouts
Erasmus Mundus Master in Complex Systems Science
University of Warwick

Supervisor: Todd Young

June 2015

Abstract

In this report we describe the work we have done for our M2 project in the Erasmus Mundus Masters in Complex Systems Science. We have studied a model of the cell cycle. The model was conceived to explain experimental observations of respiratory oscillations in yeast cultures. An essential feature of the model is that a negative feedback mechanism robustly produces clustering. We explain the model and give a brief overview of previously obtained results. Then we discuss the case of two unequal clusters, and elaborate on why a method based on reduction does not provide accurate results. We then describe a phenomenon which we call decoupling and can be observed for certain parameter values. We offer some perspectives on our results and outline directions for further investigation. Throughout we illustrate with results obtained from numerical simulations. The appendix contains some unfinished work and supplementary material.

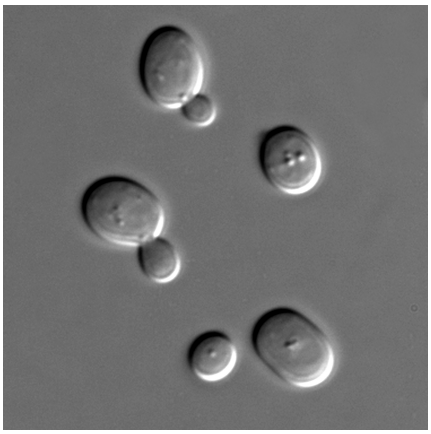
1 Introduction

1.1 Synchronization and clustering

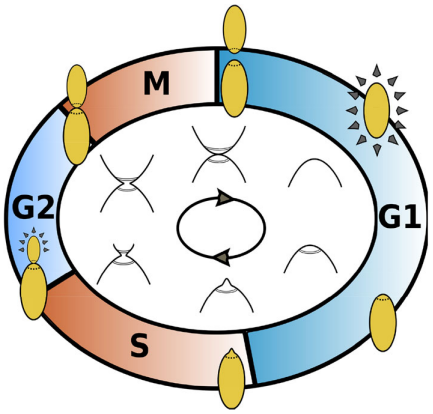
Synchronization is an often observed and well studied phenomenon. Illustrous examples in nature include the simultaneous flashing of fireflies or chirping of crickets (Strogatz, 2003). Mathematically, this is usually modeled by a system of coupled oscillators. A famous example is the Kuramoto model of coupled oscillators (Kuramoto, 1984; Strogatz, 2000). In the classic Kuramoto model, the interaction strength between oscillators with phase θ_1 and θ_2 is proportional to $\sin(\theta_1 - \theta_2)$. Depending on the strength of the coupling, synchronization may occur. This model has been extensively studied, and allows for a large number of analytic results.

A phenomenon related to synchronization is clustering, which is the topic of the present project. In a clustered state, the oscillators are synchronized in groups, or clusters. They have the same phase as the other oscillators in their cluster, but the phase differs from one cluster to another. Clustering has been studied less often than synchronization, and is less observed in nature, although there are some examples. A modified version of the Kuramoto model can produce clustering behaviour (Golomb et al., 1992; Okuda, 1993). This happens when instead of the sine function, higher order harmonics are included in the coupling function.

Other instances of clustering, however relatively sparse, can be found in a wide range of domains. Examples are clustering in networks of firing neurons (Kilpatrick and Ermentrout, 2011), or in experiments with elektrochemical oscillators (Taylor et al., 2011). In a paper by Paley et al. (2007), clustering is seen and controlled in models for the collective motion of schools of fish. Mauroy and Sepulchre (2008) have proven that clustering occurs in a model of integrate-and-fire oscillators, similar to what Mirollo and Strogatz (1990) have done for synchronization. Clustering appears in some models



(a) Microscope picture of yeast. Budded cells are clearly visible. Source: *"S cerevisiae under DIC microscopy" by Masur - Own work. Licensed under Public Domain via Wikimedia Commons*



(b) Schematic representation of the yeast cell cycle. Taken from (Young et al., 2012).

of opinion dynamics (Hegselmann and Krause, 2002), and clustering of coupled oscillators has even been proposed as the basis for an algorithm for classifying data (Rhouma and Frigui, 2001), where similarities between data points dictate coupling strengths.

Another example where synchronization and clustering occur is *yeast autonomic oscillations*, which will be explained in the next section.

1.2 Yeast, the cell cycle and oscillations

Yeast is an often studied organism in biology. Being a simple eukaryotic cell, it offers a model for more complicated animal and plant cells, and by studying it we hope to obtain some insight in the structure of these more complicated cells. Yeast cultures can be grown relatively easily in the laboratory. In addition yeast is of industrial and economic importance. It has been used for a long time in the production of food and beer, for example. The particular yeast species we are concerned with is *Saccharomyces cerevisiae*, or budding yeast (see also Figure 1a). More information on the importance of this organism in the scientific study of cells can be found in the book by Feldmann (2011).

The cell cycle is a very fundamental concept in nature. In order for an organism to grow and survive, its cells must continuously replicate. They must grow, copy their chromosomes and divide. There is a whole range of biochemical and genetic mechanisms involved in the cell cycle. An introduction can be found in the book by Murray and Hunt (1993). We will very briefly describe the concepts that are important for the understanding of the model that we have studied. Figure 1b shows a schematic representation of the cell cycle of budding yeast. The cell cycle is divided into different phases. The final phase is mitosis, at the end of which the cell divides. Budding yeast is special in the sense that division is asymmetric and procedes by budding: the cell grows a bud, which will later be the daughter cell, and during mitosis the connection between mother and daughter narrows and the daughter cell is separated from the mother. The bud leaves a distinct mark on the cell. This aids in experiments: it gives an indication of the position of a cell in the cell cycle. Figure 1a shows yeast cells, of which some are budded. The other important phase is the S-phase where the chromosomes are replicated. Mitosis and S-phase are separated by two gap phases $G1$ and $G2$.

A topic of particular interest is oscillations in yeast cultures. Experiments have shown that under certain conditions, variables such as dissolved oxygen, carbon dioxide or pH, show oscillatory behaviour. These oscillations are self-sustained and can be observed for extended periods of time in the laboratory. Different kinds of oscillations are observed, and they are usually classified according to the observable under study (such as oxygen), and the period, which can range in order of magnitude from minutes to hours. Overviews are given by Richard (2003) and Patnaik (2003). Oscillations where dissolved oxygen level is varying are usually called respiratory or metabolic oscillations, and they should be distinguished from glycolytic oscillations, which is another importand kind and typically have a higher frequency. Metabolic oscillations were thought independent of the cell cycle for a long time, since the

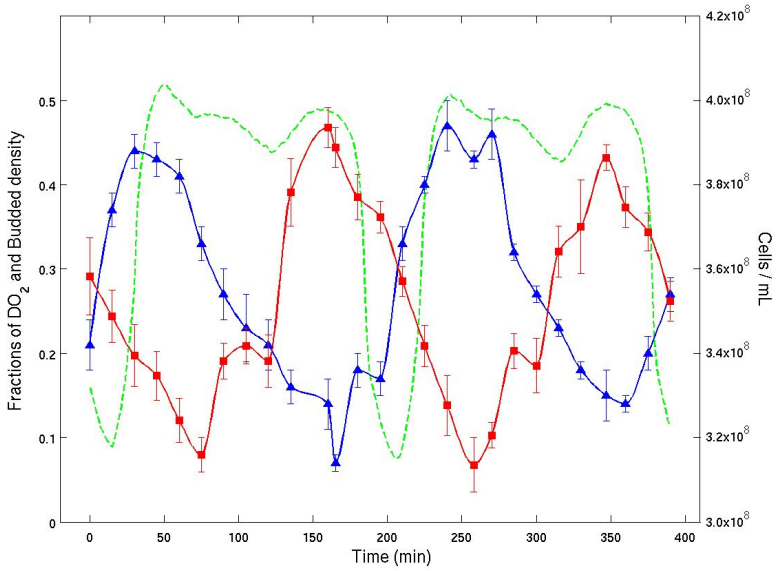


Figure 1: Figure taken from (Boczko et al., 2010), showing the dissolved oxygen (green), bud index (blue) and cell density (red).

period of the oscillations does not match the period of the cell cycle. However, a relation between the two has been noted in for example genetic expression data (Klevecz et al., 2004). A mechanism for linking the two was proposed by Robertson et al. (2008) and Boczko et al. (2010). In these papers, clustering of cells in the cell cycle was proposed as a mechanism that could explain the difference in periods. Figure 1 shows the results of an experiment conducted to test the hypothesis of clustering (Stowers et al., 2008; Boczko et al., 2010; Stowers et al., 2011). The oxygen level oscillates, as well as the bud index. The bud index is the fraction of cells budded, and gives an indication for the position of cells in the cell cycle. The bud growth commences in the S phase of the cycle. The bud index in this experiment rises periodically up to 50%, suggesting that half of the cells go into S-phase at the same time. The period of these oscillations is about half of the length of the cell cycle, which indicates that there are two clusters of cells proceeding through the cell cycle. This observation is the basis for the model studied in this project.

The study of oscillations in yeast is important since it can contribute to our knowledge of the coordination between cell cycle and cell metabolism, and for industrial purposes where oscillations are often undesirable. Oscillations can also be relevant in context of biological clocks (Lloyd et al., 2008).

As in most areas of science, experimental studies are complemented by mathematical modeling. There have been quite a few models that try to explain oscillations from yeast. Some of these models take into account a lot of biochemical variables, an example is given by Wolf et al. (2001). Other models discard the biological and chemical details, and focus on modeling basic mechanisms that might explain observed phenomena. The model we have studied is an example thereof, as will be explained in the next section. Patnaik (2003) gives an overview of some other models.

1.3 The present model

1.3.1 Explanation of the model

In this section we explain the model we studied, and review some of the important concepts. Precursors to the current model were investigated by Boczko et al. (2010) and the main formulation is given in a paper by Young et al. (2012). The cell cycle is modeled abstractly as a circle. The position of a cell in its cycle is denoted by its location on the circle. When a cell hits 1, it returns to zero. This models the biology of cell division, under the assumption that we are looking at a collection of a lot of cells where the total population is constant, and where on average each cell has one descendant. This assumption

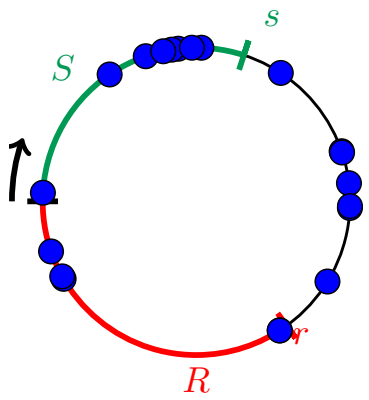


Figure 2: Possible configuration for the model. The blue dots are the cells, which move counterclockwise on the circle. The signalling and responsive regions are $[0, s)$ and $[r, 1)$.

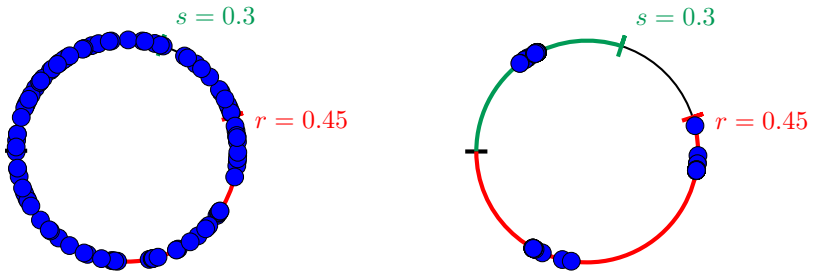


Figure 3: Left: random initial condition. Right: the system after at $T = 100$. Three clusters are formed clearly (although the simulation has not fully converged).

can be justified by the fact that in a laboratory experiment, there is outflow from the bioreactor which compensates for the growth of new cells.

A distinctive feature of the model is that cells in one region of the cell cycle (denoted S for signalling) influence the cells in another region (R for responsive). The signalling region comes after the responsive region, and is thought to correspond to the actual S region in the cell cycle where DNA is synthesized. The feedback may be exerted through a chemical that is excreted by cells in this phase of the cycle. The progression of the cells in the cell cycle is modeled by the following system of differential equations, where x_i denotes the position of the ith cell:

$$\frac{dx_i}{dt} = \begin{cases} 1 & \text{if } x_i \notin R \\ 1 + f(I) & \text{if } x_i \in R, \end{cases} \tag{1}$$

where I is the fraction of cells in the signalling region and f is a feedback function. The regions S and R will be taken $[0, s)$ and $[r, 1)$ respectively, in all that follows. Since $0 \sim 1$, the regions are adjacent. A possible configuration is illustrated in Figure 2.

There are some conditions on the feedback function: it has to be monotone and satisfy $f(0) = 0$. If f is positive, cells in the responsive region will be sped up by cells in the signalling region. This case will be referred to as positive feedback. For negative f, the cells are slowed down. This case is referred to as negative feedback. The most important characteristic of this model is that it produces clusters for negative feedback. Depending on the parameters s and r, the state where cells are grouped into k clusters ($k \geq 2$) is attracting. This is illustrated in Figure 3. For positive feedback on the other hand, the state where all cells are grouped into one big cluster (synchronization) is stable and attracting. The model has been studied an extended in a series of papers. We will give a brief overview of them in the next section.

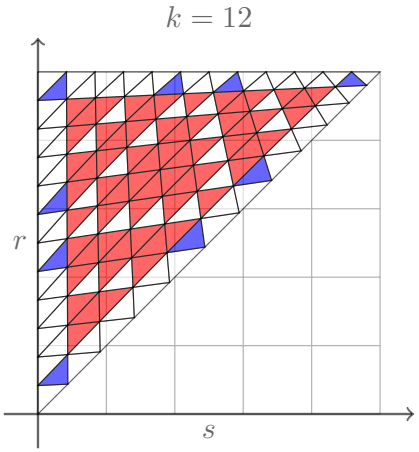


Figure 4: Isosequential regions for 12 clusters and negative feedback. Blue regions are stable, red unstable and white neutrally stable. Along the edges, blue triangles correspond with numbers that are coprime with 12.

1.3.2 Overview of earlier results

As mentioned above, a slightly different version of the model was studied by Boczko et al. (2010). In this paper clustering was observed for so-called advancing and blocking models. In these models, cells were either advanced forward with a jump or blocked at a position when the number of cells in the signalling region exceeds a threshold. The model in the form as we study it was put forward and formalized by Young et al. (2012). In this paper the basic behaviour of the model was outlined. The main tool for studying the model, the Poincaré map, was brought into play and the two-cluster case was studied in detail.

In a paper by Breitsch et al. (2014), the stability of k-cyclic solutions is studied for arbitrary k. A k-cyclic solution is a special kind of periodic solution of the system, which has more symmetry. For a k-cyclic solution, the stability can be characterized by the order of events, another important concept in the study of the model. An event happens when a cell crosses the boundary of one of the regions. For a given k-cyclic solution, the (s, r) parameter space can be subdivided into so-called isosequential regions. In each of this regions, events occur in the same order. The order of events ultimately determines the Jacobian matrix of the return map, and therefore determines the stability of the solution. The regions where a k-cyclic solution is stable, unstable or neutrally stable have a lot of structure (see for example Figure 4). Important results are that solutions with a lot of clusters are stable near the edges of parameter space and that there are a lot of regions with bistability. Lower numbers of clusters such as 2 or 3 are stable in a larger part of the parameter space. The paper ends with some conjectures about the structure of the parameter space (see also Appendix B.1 in this context).

Two other papers have studied the model with a gap between the R and S region (Gong et al., 2014a) and with noise (Gong et al., 2014b). The former model represents a delay in the signalling by a gap between the regions. The clustering behaviour is conserved, and the authors conclude that a gap might enhance the stability of the clustered solutions. In the latter article, the cells are subject to different kind of noise. Again, clustering behaviour is conserved when the noise is not too large. When the noise grows, however, the clustering breaks down, and a order-disorder phase transition occurs which resembles an Andronov-Hopf bifurcation. From this, the authors derive an approximate magnitude for the coupling strength in real biological systems, which needs to be quite large to overcome the cell's tendency to disperse due to noise.

A final paper in the series (Buckalew, 2014) adds an explicit equation for the chemical agent that does the signalling. It introduces interesting new dynamics, where clustered solutions become stable when the population density is high enough (which is related to the biological concept of quorum sensing).

A nonpublished work, which nonetheless deserves mentioning, is the Master's thesis of Rob Wesselink (2013). A slightly different version of the model is studied, which is more general but is only concerned with two clusters. The generality lies in the fact that a sensitivity function is introduced, which allows

for the cell experiencing different amounts of feedback depending on its position in the cycle. The analysis is rigourously mathematical and some interesting propositions, such as a reflection principle, are proven. The clusters studied in that work have weights, something we considered for our project and will be explained below.

2 Project: aim and methods

The principal goal of the project was to address some questions which have been left unanswered in previous papers. The most important question was what happens when clusters do not have the same size. The indistinguishability of the clusters plays a key role in most of the obtained theoretical results. This assumption allows to reduce an N-dimensional system of equations, where N is very large, to a k-dimensional system, where k is the amount of clusters, usually a small integer. While exploring this question and looking at the results of simulations, we made some other interesting observations, such as decoupling, which will be described below.

The main tool we used was numerical simulation. We relied heavily on a combination of Python and C code to do the work. Two C programs were used. One program does the simulation of the dynamical system, the other one implements a hierarchical clustering algorithm. The clustering algorithm groups cells together when their distance is small enough, and this allows to determine how many clusters are formed without needing to visually inspect data or plots. With this programs we were able to run many simulations and do statistical analyses of the results. Python was used to control the execution of the C programs, format the resulting data and visualize the results. In total we did about 30 numerical experiments. Each experiment was run on the Warwick Cluster of Workstations (CoW), since they were computationally demanding. In the appendix we list all the experiments we carried out. Not all of them are mentioned in the main text since they did not all lead to interesting results.

In addition to numerical simulation, we carried out some pen-and-paper work in the best mathematical tradition. To help our intuition, some symbolic computations were carried out using SymPy, a symbolic computation package for Python. In the next sections the results are outlined.

3 Dynamics of two unequal clusters

3.1 Poincaré map and cyclic solutions

Young et al. (2012) describe the dynamics for 2 equal clusters. The work in that paper is an illustration of a technique that proves to be very useful in the study of this system. The amount of cells in a real bioreactor can be of the order $N = 10^{10}$. This is inconvenient in computations to say the least, but one can reduce the dimensions greatly by assuming that the cells are divided into k equal clusters. Since cells that are synchronized at one point in time will stay synchronized for all future times, we only need to look at k equations, for the k clusters. These equations are exactly the same equations as for the individual cells, since in the feedback function only the fraction of cells is important. Instead of looking at the continous time dynamics, the Poincaré map P is studied. We can take a Poincaré section at $x_0 = 0$, and look at the positions of other clusters every time x_0 crosses 0. An important observation is the fact that, since the k clusters are identical, the Poincaré map can be factorized. We have $P = F^k$, where F is the mapping defined by

$$F : (x_1(0), x_2(0), \cdots, x_k(0)) \mapsto (x_0(t), x_1(t), \cdots, x_{k-1}(t)), \tag{2}$$

where t is the time at which the k-th cluster hits one. We can now relabel the indices and look at the map F again, and so forth until, after k times, x_0 hits one and we obtain the Poincaré map. The stability properties of F and P are the same. In stability results, usually the focus is on a specific type of solutions called k-cyclic solutions. These solutions are fixed points of the map F (and therefore, also of P). In a cyclic solution, we have that $x_i(d) = x_{i+1}(0)$ for some time d, i.e. every cluster takes the position of the next cluster after a certain time.

In the case of two equal clusters, the Poincaré map and F are mappings from the unit interval to itself (with 0 and 1 identified). This makes them easy to visualize.

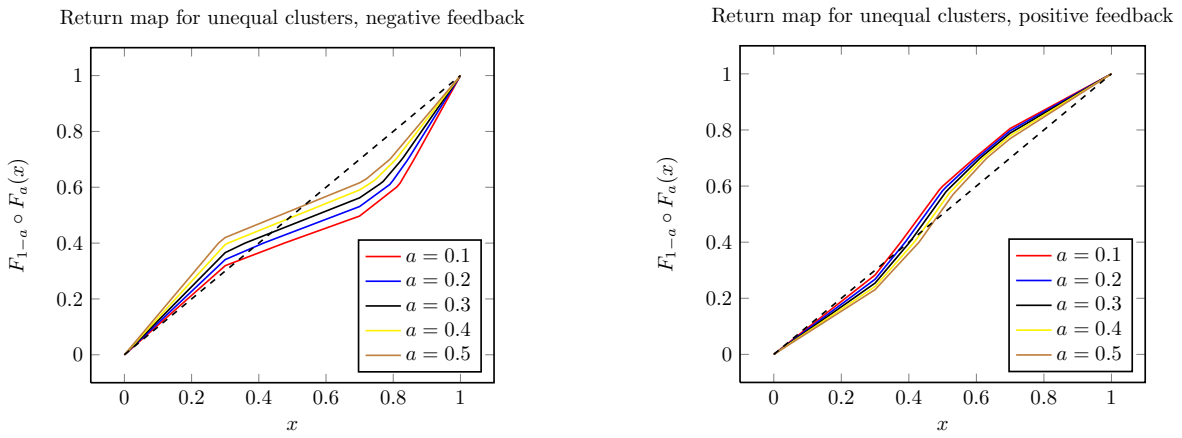


Figure 5: Return map for unequal clusters, a is the weight of the cluster starting at 0. Left: negative feedback, right: positive feedback. Parameters: $s = 0.5$, $r = 0.7$.

We have done a similar analysis for 2 unequal clusters. We look at the case where the cluster x_0 has weight a (denoting the fraction of cells in that cluster), and x_1 weight $1-a$. We can define the map F_a as follows: $F_a(x) = x_0(t)$, where t is the time it takes for x_1 to hit 1, starting at position x and when the cluster with weight a starts at 0. The full Poincaré map is then $F_{1-a} \circ F_a$. In the paper by Young et al. (2012), explicit formulae are given for the mapping F, which can be adapted to take the weight into account. We will restrict us to a summary and discussion of the results. A distinction needs to be made depending on the value of $r - s$. If $r - s > 1/2$, the gap between S and R is so large that it is possible that the two clusters never interact. They are called *isolated*. This is reflected on the return map: there is an interval of neutrally stable fixed points. This behaviour carries over to the case of unequal clusters. If $r - s < 1/2$ however, the concept of a cyclic solution needs to be abandoned. A cyclic solution would need to be a fixed point of both F_a and F_{1-a}, which is only possible when $a = 1/2$, so for when the clusters are equal.

Figure 5 shows the return maps for different values of the weight a. We can see that there is always a fixed point, which is stable for negative feedback and unstable for positive feedback, analogous to the case of two equal clusters.

3.2 Numerical experiments

The previous result seems counterintuitive: it implies that for any distribution of cells, the unequal distribution is stable. This contradicts the results from experiments, where we always see two approximately equal clusters in simulations. Figure 6 is an illustration of this. Differences in cluster sizes are always relatively low, although they seem to increase when noise is added, when N becomes larger, and when there is bistability between the 2 cluster solution and a higher number of clusters. This suggests that the reduction method is not adequate here, even though it gives good results for equal clusters. We note that in experiments, similar results are obtained for more than 2 clusters: for parameters where the k cluster solution is stable, the clusters have approximately equal sizes.

We performed two other experiments to check whether an equal distribution is always attracting, even when the system is initialized in two unequal clusters. For these experiments, we determined the initial condition for unequal clusters from a simulation with two weighted clusters. We ran this simulation until convergence to the periodic state, and took the final positions as initial positions for a simulation with N cells. For different weights $a \leq 0.5$, we initialized Na cells at the position of the small cluster and $N(1-a)$ cells at the position of the larger cluster.

In the first experiment, we looked at three different kinds of perturbations. We respectively perturbed all cells, only the cells in the small cluster and only those in the large cluster. When perturbing all or only the large cluster cells, the final distribution was half-half. When only the small cluster's cells were perturbed, the unequal distribution was conserved. This shows that cells from the large cluster move to the smaller cluster, until an equal distribution is attained. We checked this more in detail in

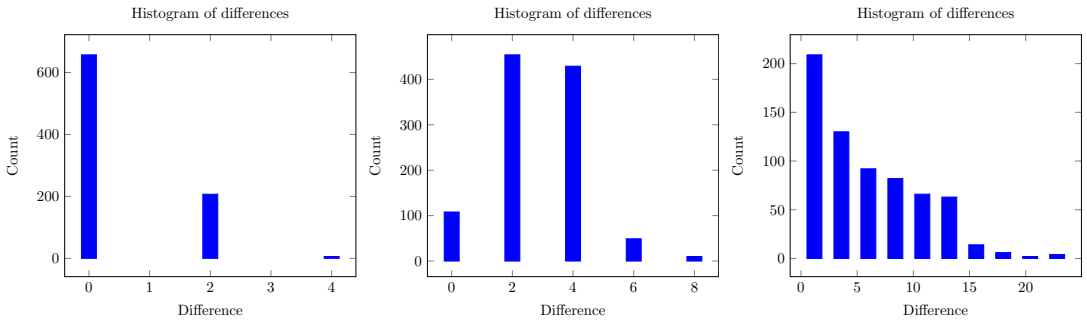


Figure 6: Histograms of the difference between largest and smallest cluster. Left: $s = 0.4$, $r = 0.65$, two is stable. Middle, same parameter values but with noise. Right: $s = 0.64$, $r = 0.67$, a region with bistability between 2 and 7 clusters. We took $N = 1000$. The clusters are almost always of near equal size, but larger differences are observed with noise and in a region of bistability.

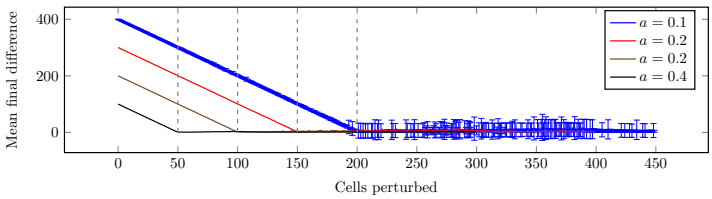


Figure 7: Results of the experiment where we perturbed one cell at a time from the largest cluster. The grey lines correspond to the amount of cells needed to move from the large to the small cluster in order to make the distribution equal. The y-axis denotes the final difference between the two clusters. It is clear that the cells tend to move from the large to the small cluster to equalize the distribution. The average is over 100 simulation, the error bars are one standard deviation. The error bars are only shown for $a = 0.1$, for clarity. They are very similar for the other curves.

a second experiment. In this experiment, we perturbed one cell at a time from the larger cluster, and looked at the final configuration. An illustration of the result is shown in Figure 7. We can deduce that each perturbed cell from the large cluster ends up in the small cluster, until $N/2 - Na$ cells are perturbed, which is the number needed to make the distribution equal. Above that, an approximately equal distribution is always seen, but there can be some deviation from the exact half-half distribution (visualized by the error bars in the figure).

The mechanism by which the cells move out from the large cluster can be the following. The cells in the large cluster are spread out due to the perturbation, and when the cluster passes from R to S, the cells that are lagging behind undergo feedback from the majority of cells in the large cluster. This pushes them farther away from the large cluster. This is not completely mitigated by feedback from the small cluster at a later time, exactly because this cluster is smaller and thus has less influence than the big cluster. This happens each time the R-S boundary is crossed and the cells are thus pushed farther from the large cluster until they are absorbed by the smaller one.

3.3 Conclusion on two unequal clusters

We can conclude that reducing the N-dimensional system to a 2 dimensional system does not give accurate results when the clusters have unequal size. For equal clusters however, the stability of the reduced system corresponds with the stability of the system in full phase space. We implicitly use this, when looking at the stability triangles to pick suitable parameter values. The fact that it holds is non-trivial, and has been proved recently (Moses, 2015).

The reason why the method fails here is the fact that looking at 2 clusters corresponds to restricting the system to a 2-dimensional subspace of the N-dimensional phase space. The positions of all cells are restricted, but also the perturbations. The reduction shows that, when restricted to perturbations in the subspace (where the whole cluster is perturbed), the unequal solution is stable. The experiments however show that it is unstable to perturbations which are transversal to this subspace (where

individual cells can be perturbed out of the cluster). Originally we thought that a bifurcation might occur for a certain value of a, below which the unequal clusters are unstable. This probably happens in the full phase space, but in the 2-dimensional subspace it does not.

A sensible thing to ask is then, how can we obtain an accurate result on the two unequal cluster system? One way is to look at linear stability of a state with unequal clusters in the full phase space. This requires looking at the full Jacobian, which will be an $N \times N$ matrix. But since we are interested in the state where a fraction a of cells are in the same position, and the $1 - a$ others in another, the Jacobian will have a specific structure. This could be exploited to obtain results on the spectrum of the Jacobian and ultimately on the stability of the state with unequal clusters.

Another option is to use a formulation with partial differential equations, where densities are used instead of discrete cells and clusters. This formulation requires the use of new techniques and might be difficult to tackle.

We finally note that the same problem, of stability in a clustered subspace versus the full space, has been considered by Golomb et al. (1992) and Okuda (1993) for a generalization of the Kuramoto model. It might be fruitful to adapt their approach to our system.

4 Decoupling along the diagonal

In the previous section we have seen that, in regions were k clusters are stable, the resulting clusters have near equal sizes. There are however some situations in which this behaviour breaks down. We have reproduced the stability triangles from the article by Breitsch et al. (2014), and took a next step: to overlay triangles for different k, and see what this implies about the corresponding k cluster solutions. In regions of bi- or multistability, we observe different numbers of clusters, although not all equally frequently. For example: in one experiment we counted how many times we observed two and three clusters for $s = 0.5$, $r = 0.8$. Without noise it was 67 times 2 clusters and 7 times 3 clusters. With noise the distribution was 90 to 10. This suggests the basin of attraction for 2 clusters is greater.

We observed something interesting when $k = p \times q$, and k clusters is a neutrally stable solution while p is stable. In this case, we often observe k clusters, but grouped in p groups. We call this *decoupling*. Figure 8 shows an illustration, and more pictures can be found in Appendix A. This decoupling also happens when noise is added and with different feedback functions, showing that it is quite robust. We have performed simulations for different regions of parameter space to see when the decoupling happens. The decoupling is only seen in the regions that lie along the diagonal. This is unexpected, since the stability triangles themselves are usually symmetric, which suggests that the behaviour along the three edges is similar. The diagonal thus shows special behaviour.

Most of the stability triangles show all symmetries of the triangle: rotations and mirroring (along the antidiagonal). Our present result suggest that only the mirroring symmetry is relevant. The stability triangles for $k = 9$ and $k = 14$ (see Appendix A), curiously, break the other symmetries. We have looked into this symmetry, but haven't advanced far enough to give a proof of it.

When decoupling occurs, a wide range of cluster sizes is seen, in sharp contrast with the almost equal cluster sizes that are usually observed. Figure 9 shows a comparison of histograms for two 4 cluster solutions, one decoupled and one non-decoupled. It is clear that the differences are much larger in the case of decoupling.

When a k cluster solution decouples into p groups of q clusters, we can see this as a superposition of q times a p-cluster solution. In each of the p groups, the same distribution of cluster sizes within this group is observed, and the total amount of cells in each of the p groups is approximately equal. There is a proposition in (Breitsch et al., 2014) which is related to this. Proposition 3.7 says that a weighted average of cyclic solutions that follow the same number of events is also a cyclic solution. This seems to be relevant to the observations we make here. To clarify this, see for example Figure 10. Here we see 4 clusters, in two groups of two. We can take the two large clusters as one 2 cluster solution, and the two small ones as another solution. Their superposition is thus a new solution, and the fact that it doesn't collapse onto two clusters might be because the four cluster solution is not repelling, but neutrally stable. This is largely a conjecture, though.

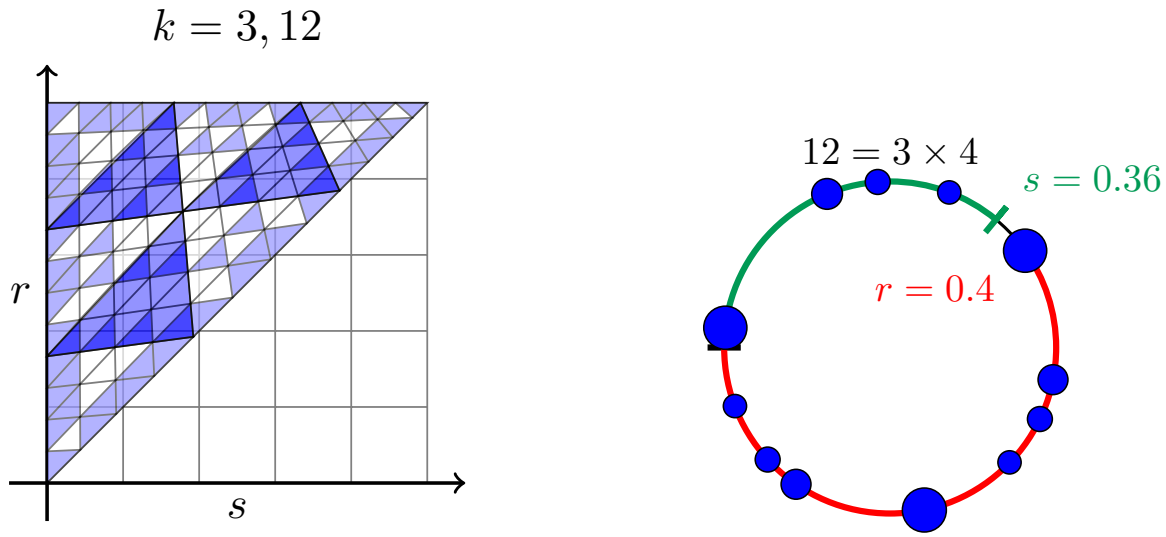


Figure 8: Left: overlay of stability regions for $k = 3$ and $k = 12$. Dark blue are regions where 3 is stable, lighter blue where 12 is neutrally stable. Right: clusters obtained for a simulation with parameter values from one of the triangles along the diagonal where 3 is stable and 12 unstable. It can be seen that the 12 clusters decouple into 3 groups of 4 clusters. The size of the dots is proportional to the amount of cells in the cluster.

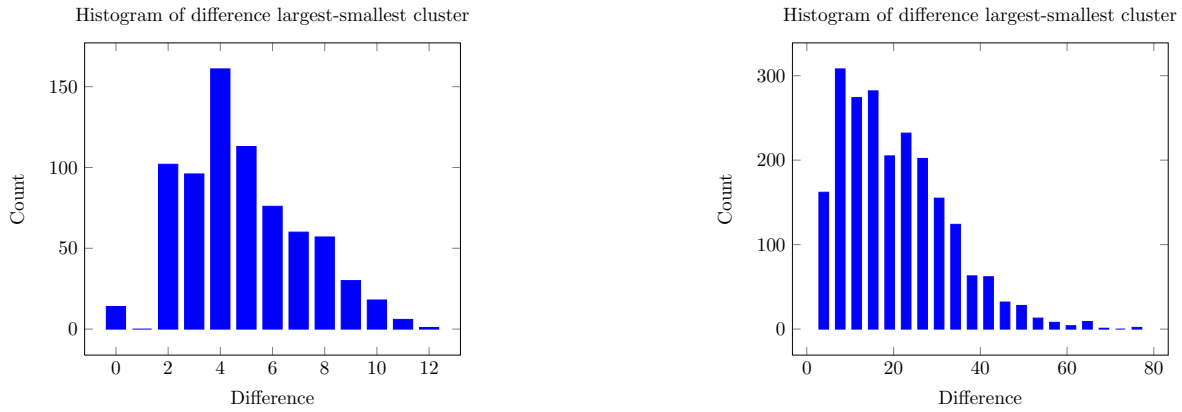


Figure 9: Histogram for the difference between largest and smallest cluster observed in 4 cluster solutions. Left: $s = 0.2$, $r = 0.35$, four clusters is stable. Right: $s = 0.45$, $r = 0.6$, four decouples into two times two.

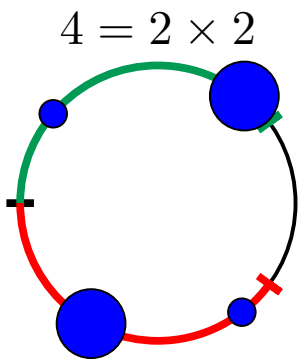


Figure 10: Four cluster solution decoupled into two groups of two. The size of the dots is proportional the the amount of cells in the cluster.

5 Discussion

We have discussed some problems and perspectives on the two unequal cluster case in Section 3.3 and on decoupling in the previous one. We now finish with some general remarks. It is clear that the study of this model, however simple it may seem at first, is not yet complete. The two unequal cluster discussion shows that there is need for another method to study this. An equal division of cells among clusters seems to be in some sense the most stable state, even though small deviations are often observed. It would be interesting to quantify how large these deviations can be. A mathematical explanation of the decoupling phenomenon is also worth investigating. It is remarkable that the decoupling only occurs in regions along the diagonal, for s and r close to each other. The diagonal therefore has a special role in the model, even though the regions of stability seem to satisfy more symmetries (but see the case of 9 and 14 clusters). This symmetry can be further investigated. We also refer to the final paragraphs of (Breitsch et al., 2014) for some conjectures and open problems on the stability regions.

In this project, we checked if decoupling happens when noise is added and when a different feedback function is used. Application to the gap model (Gong et al., 2014a) and the mediated model (Buckalew, 2014) can be checked (some preliminary simulations with the gap model show that decoupling can happen here, as well, but this is expected to depend on the size of the gap).

This model is very simple, and the mechanism that leads to clustering is very straightforward. We would expect to see more examples of clustering in nature. It would be of interest to examine if the model can be applied to other phenomena in biology or other domains. After having shown that clusters are possible in the model, a next step could also be to investigate the possibility to control the amount of clusters formed, for example by forcing the model. We imagine this could have applications in experiments with yeast in bioreactors, were pulses of chemicals can be added.

To conclude, this model is a good example of how simple mechanisms can lead to complicated and intriguing behaviour. It shows how mathematics and numerical simulation can be combined to obtain more insight in a biological problem.

6 Acknowledgements

I would like to thank Todd Young his patience and supervision. I am grateful to everyone involved in the Erasmus Mundus programme for providing me with wonderful opportunities over the last two years.

References

Boczko, E. M., Gedeon, T., Stowers, C. C., and Young, T. R. (2010). ODE, RDE and SDE models of cell cycle dynamics and clustering in yeast. *Journal of Biological Dynamics*, 4(4):328–345.

Breitsch, N., Moses, G., Boczko, E., and Young, T. (2014). Cell cycle dynamics: clustering is universal in negative feedback systems. *J. Math. Biol.*, pages 1–25.

Buckalew, R. L. (2014). Cell cycle clustering and quorum sensing in a response / signaling mediated feedback model. *Discrete and Continuous Dynamical Systems - Series B*, 19(4):867–881.

Feldmann, H. (2011). *Yeast: Molecular and Cell Biology*. John Wiley & Sons.

Golomb, D., Hansel, D., Shraiman, B., and Sompolinsky, H. (1992). Clustering in globally coupled phase oscillators. *Phys. Rev. A*, 45(6):3516–3530.

Gong, X., Buckalew, R., Young, T., and Boczko, E. (2014a). Cell cycle dynamics in a response/signalling feedback system with a gap. *Journal of Biological Dynamics*, 8(1):79–98.

Gong, X., Moses, G., Neiman, A. B., and Young, T. (2014b). Noise-induced dispersion and breakup of clusters in cell cycle dynamics. *Journal of Theoretical Biology*, 355:160–169.

Hegselmann, R. and Krause, U. (2002). Opinion Dynamics and Bounded Confidence Models, Analysis and Simulation. *Journal of Artificial Societies and Social Simulation*, 5.

Kilpatrick, Z. P. and Ermentrout, B. (2011). Sparse Gamma Rhythms Arising through Clustering in Adapting Neuronal Networks. *PLoS Comput Biol*, 7(11).

Klevecz, R. R., Bolen, J., Forrest, G., and Murray, D. B. (2004). A genomewide oscillation in transcription gates DNA replication and cell cycle. *PNAS*, 101(5):1200–1205.

Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulence*. Springer, Berlin.

Lloyd, D., Murray, D. B., Klevecz, R. R., Wolf, J., and Kuriyama, H. (2008). The Ultradian Clock (˜40 min) in Yeast (Saccharomyces cerevisiae). In Lloyd, P. D. and Rossi, D. E. L., editors, *Ultradian Rhythms from Molecules to Mind*, pages 11–42. Springer Netherlands.

Mauroy, A. and Sepulchre, R. (2008). Clustering behaviors in networks of integrate-and-fire oscillators. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(3):037122.

Mirollo, R. and Strogatz, S. (1990). Synchronization of Pulse-Coupled Biological Oscillators. *SIAM J. Appl. Math.*, 50(6):1645–1662.

Moses, G. (2015). *Dynamical systems in biological modeling: clustering in the cell division cycle of yeast*. PhD thesis, Ohio University.

Murray, A. and Hunt, T. (1993). *The cell cycle: an introduction*. Oxford University Press, New York.

Okuda, K. (1993). Variety and generality of clustering in globally coupled oscillators. *Physica D: Nonlinear Phenomena*, 63(3–4):424–436.

Paley, D., Leonard, N., Sepulchre, R., Grunbaum, D., and Parrish, J. (2007). Oscillator Models and Collective Motion. *IEEE Control Systems*, 27(4):89–105.

Patnaik, P. R. (2003). Oscillatory metabolism of Saccharomyces cerevisiae: an overview of mechanisms and models. *Biotechnology Advances*, 21(3):183–192.

Rhouma, M. and Frigui, H. (2001). Self-organization of pulse-coupled oscillators with application to clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):180–195.

Richard, P. (2003). The rhythm of yeast. *FEMS Microbiology Reviews*, 27(4):547–557.

Robertson, J. B., Stowers, C. C., Boczko, E., and Johnson, C. H. (2008). Real-time luminescence monitoring of cell-cycle and respiratory oscillations in yeast. *PNAS*, 105(46):17988–17993.

Stowers, C., Robertson, B., Johnson, C., Young, T., and Boczko, E. (2008). Clustering, Communication and Environmental Oscillations in Populations of Budding Yeast (preprint).

Stowers, C., Young, T., and Boczko, E. (2011). The structure of populations of budding yeast in response to feedback. *Hypotheses in the Life Sciences*, 1(3):71–84.

Strogatz, S. (2003). *Sync: The emerging science of spontaneous order*. Hyperion.

Strogatz, S. H. (2000). From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1–4):1–20.

Taylor, A. F., Tinsley, M. R., Wang, F., and Showalter, K. (2011). Phase Clusters in Large Populations of Chemical Oscillators. *Angew. Chem. Int. Ed.*, 50(43):10161–10164.

Wesselink, R. J. (2013). Synchronization of Oscillators.

Wolf, J., Sohn, H.-Y., Heinrich, R., and Kuriyama, H. (2001). Mathematical analysis of a mechanism for autonomous metabolic oscillations in continuous culture of Saccharomyces cerevisiae. *FEBS Letters*, 499(3):230–234.

Young, T. R., Fernandez, B., Buckalew, R., Moses, G., and Boczko, E. M. (2012). Clustering in cell cycle dynamics with general response/signaling feedback. *Journal of Theoretical Biology*, 292:103–115.

A More pictures

Both the decoupling and the stability triangles have, along with the insight they provide, a certain aesthetic quality. This appendix serves to exhibit some of this.

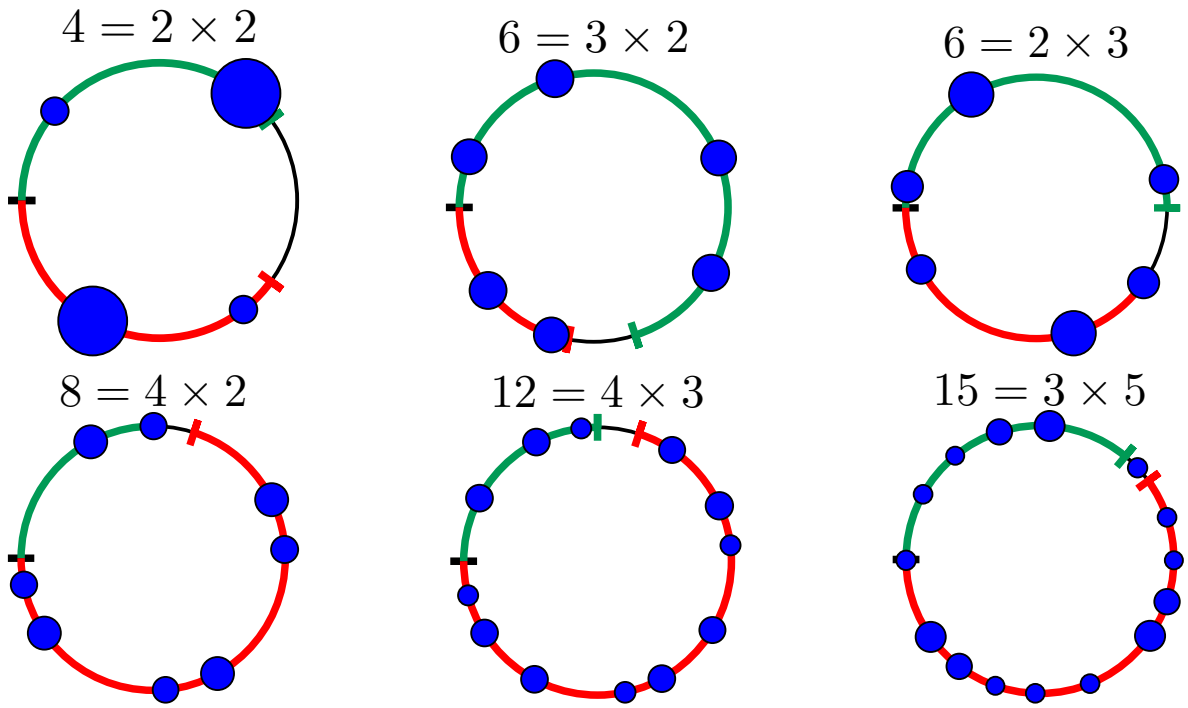


Figure 11: Results of decoupling for different values of s and r.

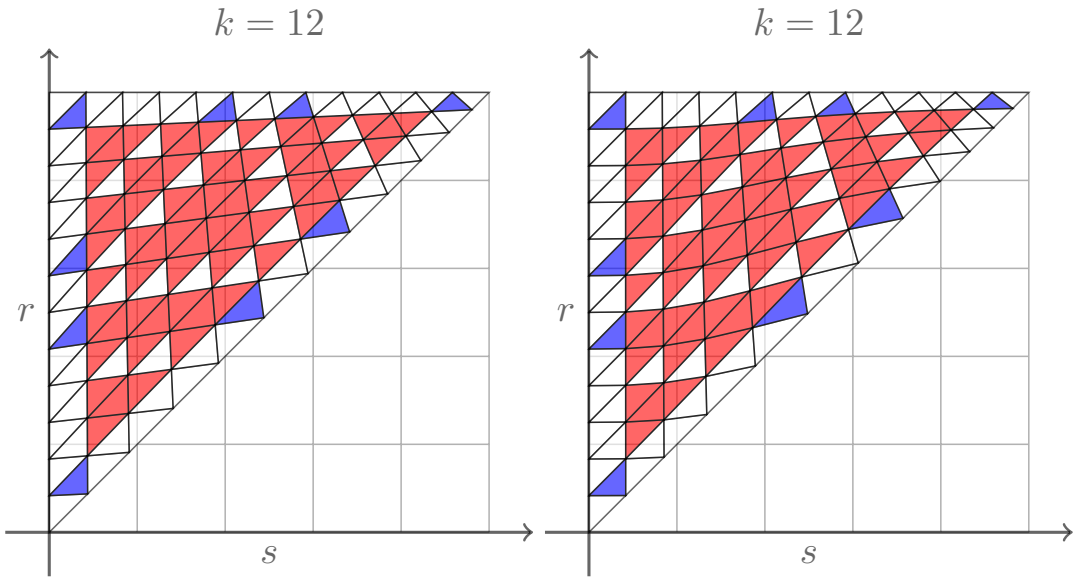


Figure 12: Stability triangles for 12 clusters, on the left for a linear feedback function, on the right for a sigmoidal feedback function. The regions themselves are still triangular, but the feedback function is clearly recognizable. The sigmoidal function used is $f(x) = -a\frac{x^c}{b+x^c}$ with $a = 0.66, b = 0.1, c = 3$.

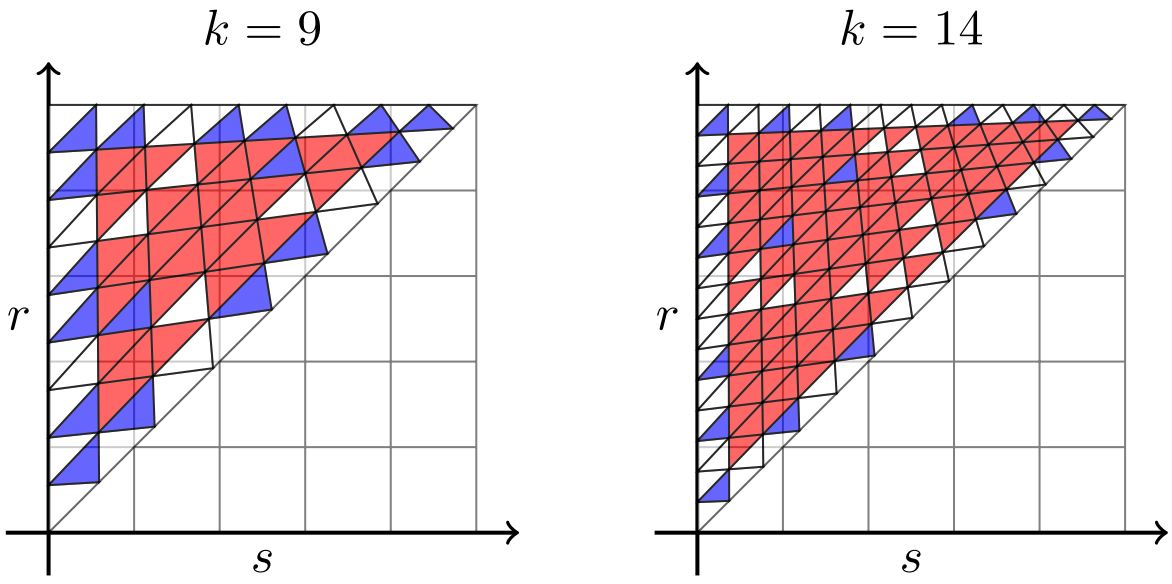


Figure 13: Regions of stability for 9 and 14 clusters. For these values, there is no complete symmetry (note the two blue triangles on the inside). For all other values, the triangles have all symmetries. See (Breitsch et al., 2014) for more pictures.

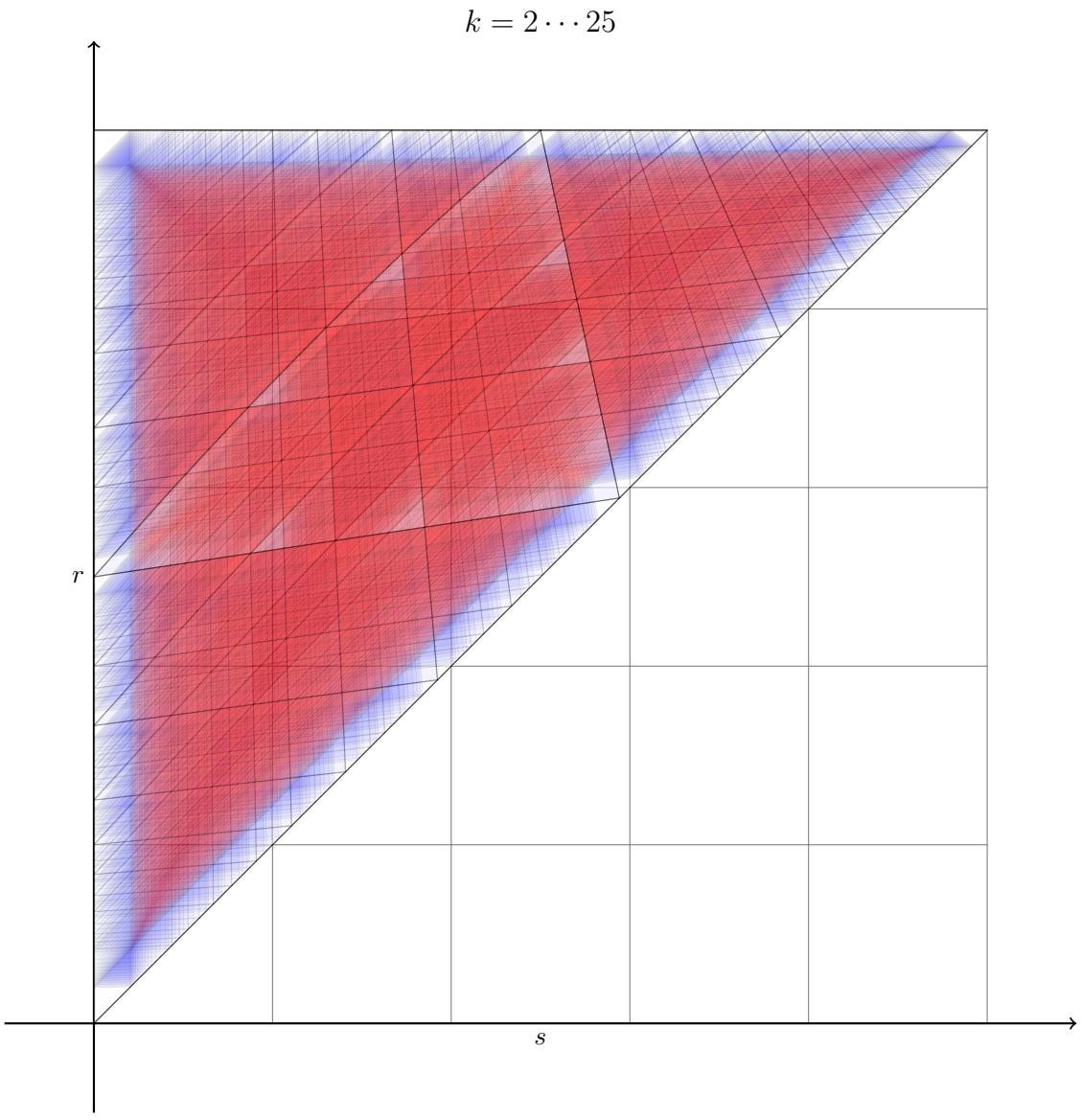


Figure 14: Overlay of stability regions for $k = 2$ to $k = 25$. Dark blue areas indicate parameter values for which a large number of k cluster solutions are stable. The large triangle in the middle is the region where 2 clusters are stable and comes out clearly.

B Partial results and observations

While working on the project, we often got sidetracked and explored some interesting things. This appendix contains some results that have not been fully developed.

B.1 Stability regions

First of all we provide a proof related to the stability triangles. This results has been recently proved in another way in G. Moses' PhD Thesis (Moses, 2015). The method we used is to obtain an analytical expression for the characteristic polynomial of the Jacobian of the factor map F, and then prove a result on the eigenvalues.

We use the terminology used by Breitsch et al. (2014), and look at a k-cyclic solution, with σ cells in the S-region and ρ cells *not in* the R-region. We take $\sigma = 1$ also assume that the first event to happen is that x_ρ reaches r. The regions in parameter space that correspond to this situation are the ones on the left edge of the triangle. The pattern of blue white can be explained as follows: if $\rho - 1$ and k are coprime, the k-cyclic is stable in this region, else it is neutrally stable. We summarize this in the following proposition.

Proposition 1. *For a k-cyclic solution with $\sigma = 1$ and x_ρ crossing r before x_σ crosses s, for negative feedback, the eigenvalues of the linear part of the factor map F are all less than 1 in absolute value if and only if* $\gcd(\rho - 1, k) = 1$.

Proof. First of all, the linear part of F has a very definite structure. We have

$$DF = \begin{pmatrix} 0 & 0 & & & & \cdots & -1 \\ 1 & 0 & & & & \cdots & -1 \\ & 1 & \ddots & & & \cdots & -1 \\ & & \ddots & 0 & & \cdots & -1 \\ & & & 1+\beta_1 & 0 & \cdots & -1 \\ & & & & \ddots & \cdots & -1 \\ & & & & & 1 & -1 \end{pmatrix}, \tag{3}$$

where the $1+\beta_1$ appears on the ρth row and $(\rho-1)$th column, and β_1 is shorthand notation for $f(1/k)$ (which is smaller than zero for negative feedback). This is a $(k-1) \times (k-1)$ matrix. We will call the characteristic polynomial of this matrix $D_\rho(\lambda)$. We will now inductively compute $D_\rho(\lambda)$. We expand $DF - \lambda I$ along the first row and obtain

$$D_\rho(\lambda) = (-\lambda) D_{\rho-1}(\lambda) + (-1)^k (1 + \beta_1), \tag{4}$$

where $D_{\rho-1}(\lambda)$ is of the same form as $D_\rho(\lambda)$, but with the $1 + \beta_1$ in the $(\rho - 1)$th row and $(\rho - 2)$th column. In addition, we have that $D_1(\lambda)$ is the determinant of the matrix

$$\begin{pmatrix} -\lambda & 0 & & \cdots & -1 \\ 1 & -\lambda & 0 & \cdots & -1 \\ 0 & 1 & & \cdots & -1 \\ & & & 1 & -1-\lambda \end{pmatrix},$$

with dimensions $(k - \rho) \times (k - \rho)$. This evaluates to $\sum_{i=0}^{k-\rho} \lambda^i$ (using a similar induction). Now we use the lemma below to get

$$D_\rho(\lambda) = \lambda^{\rho-1} \sum_{i=0}^{k-\rho} \lambda^i + (1 + \beta_1) \sum_{i=0}^{\rho-2} \lambda^i. \tag{5}$$

We now prove that the roots of this polynomial cannot be larger than 1 in absolute value, and that if $\rho - 1$ and k are coprime, there is a root with absolute value 1. For the latter part, note that the roots

of $\sum_{i=0}^{n} \lambda^i$ are the $(n+1)$th roots of unity. From basic number theory we know that if p and n are coprime, their roots of unity are distinct, and when they are not coprime, there is a root in common. Applied to the polynomial above, this shows that when $\rho - 1$ and $k - \rho + 1$ are coprime, they do not have a common root, and if not, they do have it. Note that $\gcd(\rho - 1, k - \rho - 1) = 1$ is equivalent to $\gcd(\rho - 1, k) = 1$.

To show that none of the roots can have absolute value of more than 1, we can do something analogous to a proof in the appendix of (Young et al., 2012). Assume that λ is a root and that $|\lambda| > 1$. Dividing Eq. (5) by $1 - \lambda$ and putting equal to zero, we have

$$\begin{aligned} \lambda^{\rho-1}(1 - \lambda^{k-\rho+1}) + (1 + \beta_1)(1 - \lambda^{\rho-1}) &= 0 \\ \Leftrightarrow \lambda^{\rho-1} &= \frac{1 + \beta_1}{\lambda^{k-\rho+1} + \beta_1}. \end{aligned} \tag{$\star$}$$

Now note that $\beta_1 < 0$. Using the reversed triangle inequality, we have that $|\lambda^{k-\rho+1} + \beta_1| > ||\lambda|^{k-\rho+1} + \beta_1| > |1 + \beta_1|$, where the last inequality follows since $|\lambda| > 1$, by assumption. Therefore the right member in ($\star$) has absolute value less than one, which gives a contradiction since by assumption the left hand side has absolute value greater than 1.

□

Lemma 1. *If a sequence a_k satisfies*

$$a_{k+1} = x a_k + b_k \qquad \text{and } a_1 = a, \tag{6}$$

then we have

$$a_k = a x^{k-1} + \sum_{i=1}^{k-1} b_i x^{k-1-i}. \tag{7}$$

Proof. The lemma is easily proved by straightforward induction. □

The left edge of the stability triangles is probably the easiest one, since the matrix DF has a very regular structure. It might be possible to prove similar results for other parts of the parameter space, where the structure is also regular.

We have contemplated proving an analogous result for the upper edge of the stability triangle, for $\rho = k$. The matrix DF is more complicated here though, and a more elegant way of proving the result would be to prove that the triangle exhibits mirror symmetry: the region with $\sigma' = k + 1 - \rho$ and $\rho' = k + 1 - \sigma$ has the same stability as the region with σ, ρ. One observation is that inverting the map F around the cyclic solution gives the right correspondence between σ, ρ and σ', ρ' in each region, but now the region S is the responsive and R the signalling region. We have not proceeded further with this approach.

C List of experiments

This section gives an overview of all the numerical experiments carried out, together with why they were carried out and the obtained result. If a cluster number is mentioned in bold, it means this was the most observed number of clusters. Feedback is negative and linear unless otherwise noted, and initial condition is random. In experiments where different regions are checked, the a regions lie along the diagonal while b and c are on the other edges of the stability triangle.

Remark: a small mistake was found in the application of the noise. The noise added was uniform on $[0, \sigma]$, while it should be on $[-\sigma/2, \sigma/2]$. This was only changed from experiment 26 onwards. Some checks were carried out and none of the results seemed to be influenced by this mistake.

The list below was mainly used as a personal reference and may therefore contain sloppy notation, errors, or both. Question marks denote that the exact parameter values are unknown.

Table 1: Overview of the numerical experiments.

Experiment	Description	Goal	Result
1	$s = 0.3$, $r = 0.4$, $N = 1000$,$\sigma = 0$	Initial test run	Some errors, for example sometimes one big cluster
2	?	?	Always 2 clusters observed.
3	$s = 0.4$, $r = 0.65$, $N = 1000$,$\sigma = 0$	Observe 2 clusters (stable)	Always 2 clusters observed, with low differences in size.
3b	Same as 3 but with $\sigma = 0.003$ (?)		Always 2 clusters observed.
4	$s = 0.3$, $r = 0.5$, $N = 1000$,$\sigma = 0$	Observe 3 clusters (stable)	Always 3 clusters observed, with low differences in size.
4b	Same as 4 but with $\sigma = 0.003$ (?)		Always 3 clusters observed.
5	$s = 0.2$, $r = 0.35$, $N = 1000$,$\sigma = 0$	Observe 4 clusters (stable)	Always 4 clusters observed, with low differences in size.
5b	Same as 5 but with $\sigma = 0.003$ (?)		Always 4 clusters observed.
6	$s = 0.16$, $r = 0.45$,$N = 1000$,$\sigma = 0$	Observe 5 clusters (stable, I thought)	3 clusters observed, not 5 (not the right parameter values probably)
6b	Same as 6 but with $\sigma = 0.003$ (?)	Observe 5 clusters	3 clusters observed, not 5
7	$s = 0.3$, $r = 0.4$, $N = 1000$,$\sigma = 0$	Redo of experiment 1 after errors have been corrected	3, **6** and 7 clusters observed. Decoupling of 6 into 2×3. Note: seemingly more 3 cluster solutions when N is lower, more 6 cluster solutions when N is larger. Also this experiment was redone.
8	$s = 0.3$, $r = 0.5$, $N = 500$,$\sigma = 0$	Redo of experiment 4 to observe 3 clusters, but with more runs	Always 3 clusters observed with low differences
9	Same as 8 but with $N = 900$	Check whether the fact that N is divisible by 3 changes anything	Similar results as experiment 8
10	$s = 0.3$, $r = 0.4$, $N = 1000$ (?), $\sigma = 0.003$	As experiment 7 but with noise	Same results as 7, I noted seemingly faster convergence with noise
11	$s = 0.65$, $r = 0.7$	See what happens. Values of s, r close to diagonal, we expect a lot of clusters	Almost always **12** clusters, sometimes 17 observed. Strong decoupling for 12 clusters
12	? Not on CoW	Running 2 unequal clusters with small perturbations to check stability	1 cell perturbed, seems to stay, generally 50/50 ending distribution
13	region a: $s = 0.5$,$r = 0.6$, region b: $s = 0.2$,$r = 0.6$, region c: $s = 0.4$, $r = 0.85$. $N = 1000$, $\sigma = 0.03$	Check if decoupling occurs in all regions where 4 is neutrally stable and 2 stable	Problem with output files, but region b and c give 2 cluster solutions (so only decoupling in a)

14	region a: $s = 0.45$,$r = 0.6$, region b: $s = 0.1$,$r = 0.55$, region c: $s = 0.45$, $r = 0.85$. $N = 1000$, $\sigma = 0.03$	Check for the regions where 6 is neutrally stable and 2 is stable	Didn't see 6 clusters, but 4, decoupled only in region a and not in b and c. In a: more 2 clusters than 4 clusters.
15	region a: $s = 0.55$,$r = 0.6$, region b: $s = 0.08$,$r = 0.55$, region c: $s = 0.5$, $r = 0.95$. $N = 1000$, $\sigma = 0.03$	Check for the regions where 8 is neutrally stable and 2 is stable	Region a: many solutions are found: **2**, 4,6,8,10 mainly, but also (very little) 5 and 7. Clear decoupling for 4, 6,8,10,12 (but for latter difficult to see since few occurences). Histogram for differences in 4 clusters unexpected shape: no peak but "plateau". Region b and c: almost only 2 clusters, sometimes 4,6 or 7 but very little.
16	region a: $s = 0.5$, $r = 0.8$, region aa: $s = 0.3$, $r = 0.4$, region b: $s = 0.15$, $r = 0.4$, region bb: $s = 0.15$, $r = 0.75$, region c: $s = 0.3$, $r = 0.9$, region cc: $s = 0.65$, $r = 0.9$, $N = 1000$, $\sigma = 0.03$	Check for regions where 6 is neutrally stable and 2 is stable	In regions b, bb, c, cc: only 3 cluster solutions. Region aa: clear decoupling. Region a: 3 clusters, 2 clusters, no 6 $\rightarrow$ took wrong parameters, should be $s = 0.7$ instead of $s = 0.5$
17	region a: $s = 0.72$, $r = 0.8$, region aa: $s = 0.33$, $r = 0.4$, region b: $s = 0.09$, $r = 0.37$, region bb: $s = 0.07$, $r = 0.7$, region c: $s = 0.31$, $r = 0.94$, region cc: $s = 0.65$, $r = 0.95$, $N = 1000$, $\sigma = 0.03$	Check for regions where 9 is neutrally stable and 3 is stable	Regions a and aa: decoupling of 6 and 9 clusters. Curious histogram for region a and differences of 9 clusters: seemingly uniform distribution. But sample size here is small (≈ 10)
18	region a: $s = 0.53$, $r = 0.6$, region aa: $s = 0.53$, $r = 0.7$, region aaa: $s = 0.42$, $r = 0.59$, region b: $s = 0.08$, $r = 0.53$, region bb: $s = 0.18$, $r = 0.56$, region bbb: $s = 0.18$, $r = 0.65$, region c: $= 0.49$, $r = 0.95$, region cc: $s = 0.40$, $r = 0.86$, region ccc: $s = 0.50$, $s = 0.88$, $N = 1000$, $\sigma = 0.03$	Check for regions where 10 is neutrally stable and 2 is stable	No 10 clusters observed. Decoupling in region a: $4 = 2 \times 2$

19	$s = 0.53$, $r = 0.6$. Initial perturbation of the cells: added random number uniform from $[-\epsilon/2, \epsilon/2]$ with $\epsilon = 0.03$. $N = 1000$, values of a: $0.05, 0.1, \ldots, 0.5$	Simulations, starting with two unequal (weight $a, 1-a$) clusters at the correct initial condition (obtained from simulation with two weighted clusters). Different perturbations: all cells, only cells in smaller cluster, only cells in larger cluster	Unfortunate parameter choice: a lot of other different cluster sizes are stable
20	$N = 500$, $\sigma = 0$, grid 51×51	for parameter values s, r, run simulations, check when two clusters are created and calculate the mean of the difference to see if this is different for different parameter values	Note: first tried with finer grid and more cells, but took too much time
21	Redo of experiment 19 with $s = 0.3$, $r = 0.65$	Perturb unequal clusters, see what happens	When perturbing all cells or only the cells in the large cluster, there is always equilibration to a half-half distribution (approximately). When only perturbing the smaller cluster, the sizes are conserved
22	Same as 20 with noise	See how the differences differ over parameter space and whether there is a difference with or without noise	
23	Same as 21 without perturbation but with noise $\sigma = 0.03$	Check what happens with noise	Always equilibration to half-half distribution
24	$s = 0.3$, $r = 0.65$, $N = 500$, $a = 0.05, 0.1, \ldots, 0.5$	Start with unequal clusters at correct position, perturb cell by cell from the larger cluster	The cells always move to the smaller cluster, until enough are perturbed to make the distribution half-half. When more cells are perturbed, an approximately half-half distribution is always seen
25 and 25b	Redo of 24, but $a = 0.4, 0.41, \ldots, 0.6$ and we don't go until all cells in the largest cluster are perturbed, only up to $N/2 - (1-a)N$, enough to make half-half	Get a finer result than for experiment 24, and check whether maybe the inequality is preserved when close to half-half	Same result as in 24: all cells go to the smallest cluster to get half-half
26	For 2,3: $s = 0.5$, $r = 0.8$, 2,5: $s = 0.5$, $r = 0.86$, 2,7: $s = 0.5$, $r = 0.91$, $N = 1000$, $\sigma = 0$ and $\sigma = 0.03$	Regions of bistability: see what fraction ends up in either state	For 2,3: without noise 67-7, with noise: 90 - 10. For the other: bad parameter values

26b	redo of 26 but with 2,5: $s = 0.52$, $r = 0.7$ and 2,7: $s = 0.64$, $r = 0.67$		2,3, no noise: 508 - 54, with noise: 679-44. 2,5, no noise: 701-0, with noise: 730-0. 2,7, no noise:602-13, with noise:668-2
27	$s = 0.3$, $r = 0.65$, $N = 100, 200, 500, 1000, 2000$, $\sigma = 0$ and $\sigma = 0.03$	See how the differences differ for N, with and without noise	More variance with noise, also with N (expected)
28	Same as 25, but now we perturb more cells than are needed to equilibrate to half-half		Problem with the output files, no sensible results