Core-halo instability transition in complex systems

Seth Lloyd

Department of Mechanical Engineering

Massachusetts Institute of Technology

MIT 3-160, Cambridge, MA 02139 USA

Santa Fe Institute

1399 Hyde Park Road, Santa Fe, NM 87501 USA

Abstract: This paper proves a network instability theorem. As one adds interactions between subsystems in a complex system, structured or random, a threshold of connectivity is reached beyond which the overall dynamics inevitably goes unstable. The threshold occurs at the point at which flows and interactions between subsystems ('surface' effects) overwhelm internal stabilizing dynamics ('volume' effects). The theorem is used to identify instability thresholds in systems that possess a core-halo structure, including the gravothermal catastrophe – i.e., star collapse and explosion – and the interbank payment network. The same dynamical instability gives rise both to gravitational collapse and to financial collapse.

A wide variety of work addresses the stability of complex systems made up of networks of interacting subsystems [1-5]. A key ingredient of stability is network connectivity [5]. One of the best-known results in this field is May's theorem that differential equations described by random networks undergo a transition from stable to unstable behavior at a critical value of their connectivity [4]. Networks that occur in nature are rarely random,

1

however: they typically possess complex structures related to their function [5]. This paper proves an instability theorem for structured, non-random networks and applies that theorem to networks with a core-halo structure [6-8]. Two such networks are the interbank transfer network [7-8], and the network of gravitational interactions within a star [6]. As interactions are added between core and halo, the overall system inevitably goes unstable. The theorem identifies the instability threshold, implies characteristic behaviors in the approach to instability, and suggests methods for reversing instability.

The stability to instability transition identified in this paper arises from a basic feature of coupled ordinary differential equations. Such sets of equations are ubiquitous in the mathematical modeling of dynamical systems, and can be applied to physical systems (e.g., Newtonian gravity, electrodynamics), networks of chemical reactions, biological systems (e.g., ecological models and food webs), engineered systems (feedback control), and economic and financial systems (market economies, flows of money and debt).

Interactions and instability

Consider a set of non-linear, time-dependent, ordinary differential equations over n variables:

$$\frac{d\vec{x}}{dt} = g(\vec{x}, t), \tag{1}$$

where $\vec{x} = (x_1, \ldots, x_n)$. The dynamics of a small perturbation $\Delta\vec{x}(t)$ to a solution $\vec{x}^*(t)$ obeys the linearized equation

$$\Delta\dot{\vec{x}} = \nabla g|_{\vec{x}^*}\Delta\vec{x}. \tag{2}$$

The perturbation decreases in size if

$$\frac{d}{dt}\Delta\vec{x}^\dagger\Delta\vec{x} = \Delta\vec{x}^\dagger(\nabla g^\dagger + \nabla g)\Delta\vec{x} \equiv 2\Delta\vec{x}^\dagger G\Delta\vec{x} < 0, \tag{3}$$

where $G = (\nabla g^\dagger + \nabla g)/2$ is the symmetrized Hermitian gradient evaluated at $x^*(t)$. The Hermitian part of the gradient governs exponentially increasing and decreasing behavior, while the anti-Hermitian part $\tilde{G} = (\nabla g^\dagger - \nabla g)/2$ governs oscillatory behavior. All perturbations decrease in size if and only if the Hermitian gradient G is negative definite. The threshold of instability identified in this paper occurs at the point where interactions (off-diagonal terms in ∇g and G) become sufficiently strong to make some eigenvalue of G positive, so that some perturbations grow in size. Note that the definition of stability adopted here – small perturbations decrease in size – is stronger than Lyapunov's definition of stability, which demands only that small perturbations *eventually* decrease in size [1-2].

2

The stronger definition is adopted here because the largest eigenvalue of G typically grows in proportion to system size – as shown in the supplementary material, once a large, complex system passes the stability threshold, perturbations grow large very rapidly, taking the nonlinear system to a new and unpredictable regime.

The instability threshold identified here arises from excess of interaction. Off-diagonal terms in ∇g govern interactions or flows of energy, entropy, money, etc., between subsystems of the network, and on-diagonal terms represent sources and sinks of the same quantities. The internal dynamics of subsystems correspond to diagonal blocks of the gradient matrix of the linearized equations, while flows between subsystems correspond to off-diagonal blocks. Look at the interaction between two such subsystems. Subsystem $\mathcal{A}$ consists of n_A variables, and subsystem $\mathcal{B}$ consists of $n_B \geq n_A$ variables. Let G_{AB} be the restriction of G to the subspace spanned by the $n = n_A + n_B$ variables that describe $\mathcal{A}$, $\mathcal{B}$. Assume that $\mathcal{A}$ and $\mathcal{B}$ are locally stable in the absence of interaction, and investigate how that stability changes as interactions are added. The interactions between subsystems lead to stable dynamics if all the eigenvalues of the Hermitian matrix G_{AB} are negative. Write this matrix as

$$G_{AB} = \begin{pmatrix} A & C \\ C^\dagger & B \end{pmatrix}, \tag{4}$$

where A gives the Hermitian dynamics confined to the subsystem $\mathcal{A}$, and B gives the Hermitian dynamics for $\mathcal{B}$. By the assumption of local stability, A and B are negative definite. C is an $n_A \times n_B$ matrix whose coefficients determine the strength of interactions between $\mathcal{A}$ and $\mathcal{B}$.

As more and more interactions are added, and as the strength of those interactions increase, then the interactions inevitably drive the system unstable. In particular, we have

Theorem 1: If $\operatorname{tr} C^\dagger C > \sqrt{\operatorname{tr} A^2}\sqrt{\operatorname{tr} B^2}$, then the system is unstable.

Theorem 1 is a higher dimensional generalization of the fact that a 2×2 matrix $\begin{pmatrix} \mu & \lambda \\ \bar{\lambda} & \nu \end{pmatrix}$ where $\mu, \nu < 0$, has a positive eigenvalue if $|\lambda|^2 > \mu\nu$. The proof of theorem 1 is presented in the supplementary material.

Theorem 1 implies that the interacting systems are unstable when the average magnitude squared of the terms in the destabilizing interactions is larger than the geometric mean of the average magnitude squared of the terms in the stabilizing local dynamics. Inevitably, if the strength of the stabilizing local dynamics is fixed, increasing the strength

3

of the interactions drives the system unstable beyond some threshold. Intuitively, the threshold occurs at the point where flows through the 'surface' between $\mathcal{A}$ and $\mathcal{B}$ dominate the 'volume' flows within $\mathcal{A}$ and $\mathcal{B}$. Applied to random matrices representing the interactions between two parts of a complex system, theorem 1 reproduces the results of May [4] for connection-driven instability. However, no assumptions concerning random matrix theory were required to prove the theorem – the matrices involved can be highly structured. Theorem 1 will now be used to analyze several interaction-induced instability thresholds.

Example: Feedback control

Feedback controllers are often employed to stabilize engineered systems [1-2]. In feedback control, measurements on a system are processed and fed back via actuators to control the system. The feedback signals represent interaction (i.e., off-diagonal) terms in the linearized dynamics of the controlled system, and the size of these off-diagonal terms are controlled by parameters called gains. While small gains can stabilize the system, if the system is sufficiently complex (at least three more poles than zeros), then the well-known root locus method implies that the feedback system necessarily goes unstable at some sufficiently high gain value [1-2]. When the system passes the instability threshold of theorem 1, perturbations start to grow as oscillations become increasingly undamped. Increasing interaction eventually drives the system beyond the Lyapunov threshold and oscillations grow exponentially. Feedback driven instability is a straightforward example of interaction-driven instability implied by theorem 1.

Example: Core-halo instability and the gravo-thermal catastrophe

A common type of system in the universe consists of a collection of matter, e.g. a cloud of interstellar dust, a star, or a cluster of stars in a galaxy, interacting via the gravitational force, augmented by collisions and heat production via, e.g., nuclear reactions. Such a system naturally forms itself into a dense 'core' (system $\mathcal{A}$) of strongly interacting matter at high temperature, surrounded by a less-dense 'halo' (system $\mathcal{B}$). The microscopic dynamics of such a system are complex [6]. The supplementary material derives a simple linearized version of the energy transfer dynamics between $\mathcal{A}$ and $\mathcal{B}$ in terms of macroscopic variables. The model takes the matrix form

$$\frac{d}{dt}\begin{pmatrix} T_A \\ T_B \end{pmatrix} = \begin{pmatrix} (\eta - \alpha)/C_A & \alpha/C_A \\ \alpha/C_B & (-\gamma - \alpha)/C_B \end{pmatrix}\begin{pmatrix} T_A \\ T_B \end{pmatrix}. \tag{5}$$

4

Here, T_A is the temperature of the core and C_A is its specific heat. Similarly, T_B and C_B are the temperature and specific heat of the halo. $\alpha \geq 0$ gives the linearized rate of energy transfer between core and halo as a function of their temperature difference. $\eta \geq 0$ governs energy production in the core, due, e.g., to nuclear reactions, and $\gamma \geq 0$ governs heat loss from the halo to space beyond.

The key feature of equation (4) is that the specific heat of systems whose dynamics is dominated by gravity is typically *negative*: when the hot core of tightly bound particles loses energy, the remaining particles cluster together more tightly and move *faster*. When $C_A < 0$, demanding that the system be locally stable and below the interaction-driven instability threshold requires $C_B > 0$ and $\eta > \alpha$. That is, the overall system can still be stable if the specific heat of the halo is positive, so that like an ordinary gas it grows cooler as it loses energy, and if internal heat production in the core outweighs heat loss to the halo. As the internal rate of heat production slows – for example, as the nuclear reactions inside a star burn through their fuel – the system goes unstable at the critical threshold when η becomes less than α. At this point destabilizing flows of energy from core to halo dominate the stabilizing production of energy within the core. The temperature of the core now rises exponentially in time, with exponentially increasing flows of energy from core to halo.

This accelerating, unstable flow of energy is called the gravo-thermal catastrophe [6]: from the dynamics (4) the gravo-thermal catastrophe is seen to be a straightforward instance of interaction-driven instability governed by theorem 1. Figure 1 shows the aftermath of a gravo-thermal catastrophe, the Cat's Eye nebula: the accelerating flows of energy have blown out the outer layers of the star, accentuating the core-halo structure. For a star with more than a few solar masses or for galaxy formation in the early universe, the gravo-thermal catastrophe results in gravitational collapse of the core, and the formation of a black hole. With the formation of a black hole, energy flows from core to halo cease (except for a small amount of Hawking radiation). The black hole 'freezes' the previously hot core, and reverses the direction of energy flow, sucking up matter and energy from the halo.

Example: Core-halo instability in financial collapse

Like galaxies or nebulae, the interbank payment transfer network possesses a core-halo structure [7-8], and is susceptible to interaction-driven instability. As detailed in [7], in 2007 this network consisted of over 6600 financial institutions connected by over $70,000$

daily transfers. Most of the institutions (the halo) had either few links or links whose transfers had only small volume. A small, highly connected fraction of the institutions (the core), accounted for most of the volume. On a typical day, for example, a core of 66 institutions connected by 181 links comprised 75% of the value transferred [7]. The core itself contained an 'inner' core of 25 institutions that were almost fully connected. The core-halo structure of the network is shown in figure (2). Most important for stability analysis, the interbank payment transfer network is strongly disassortative [9]: highly-connected banks do most of their business with sparsely-connected banks, and *vice versa*. The disassortative nature of the network means that there are fewer internal links within the core, and within the halo, than there are between core and halo.

Disassortative networks are known to be less stable than assortative networks with respect to mixing and link removal [9]. The results of this paper can be applied to disassortative networks corresponding to the dynamics of coupled ordinary differential equations in general, and to the interbank transfer network in particular. Theorem 1 implies that if all links correspond to terms of approximately the same size, then the dynamics of such a disassortative network are unstable (proof in supplementary material). Since links within the halo are sparse and low volume, theorem 1 implies that stability can only be obtained when the banks in the core have significantly stronger interactions with each other than with banks in the halo. That is, even though the banks in the core undergo more transactions with banks in the halo than with each other, to maintain stability, typical flows between banks in the core and and other banks in the core must be significantly larger than typical flows between banks in the core and banks in the halo. This 'hot core' requirement for stability is confirmed by the data [7] – as noted, the core contains three quarters of the flow on a given day. Only by having large exchanges with each other (e.g., by hedging) can the banks within the core overcome the disassortative nature of the network to provide stability.

The hot core requirement leaves the network vulnerable to interaction-driven instability. Theorem 1 implies that if some event causes a sudden drop in the strength of transfer rates within the core, then the whole system can go unstable. The mathematical origin of this financial instability is the same as the origin of instability in gravo-thermal collapse, where a slowing of energy production in the core drives the system unstable. The end point of the financial instability is the well-known liquidity trap, a spectacular example of which occurred during the financial crisis of 2008-2009. As with any interaction-driven

6

instability, the instability threshold is marked by increasingly large fluctuations in volume. In the case of financial instability, such large fluctuations can violate conventional risk-management assumptions, e.g., the log-normal fluctuations predicted by the Black-Scholes equation [10] become 'fat tailed' power law fluctuations instead [11-12]. The linearized analysis given here only identifies where the instability threshold occurs: future work will investigate nonlinear models of stability [13-14]. Generically, one expects that when the threshold is passed, increasing oscillations are followed by a rapid burst of exchange, followed by a transition of the complex, nonlinear system to a new regime of behavior [13]. For the bank transfer network, just as for black hole formation, in the new regime the core freezes up, and transfers drop dramatically. The common dynamical origin of financial collapse and gravitational collapse can be termed 'the black hole of finance.'

Reversing core-halo instability

Theorem 1 immediately suggests a stabilizing course of action for systems undergoing core-halo instability: reduce interaction terms to try to isolate the core from the halo, and increase local stabilizing dynamics within the core and halo. This stabilizing strategy worked for the bank exchange network in the case of the financial crisis that began in 2008. Passage through the core-halo instability threshold had already caused exchanges between core and halo (interaction terms) to drop significantly because the liquidity trap. Governments and central banks pumped money into core and halo banks individually to keep them solvent. The combination of decreased interaction and increased local stability pulled the financial system as a whole back across the core-halo instability threshold.

A similar core-halo instability apparently lies at the root of the current European sovereign debt crisis. In this case, however, the European effort to reverse the core-halo instability has consisted of efforts to reduce the value of sovereign debt, and to enforce austerity within halo countries, thereby reducing their economic activity. Theorem 1 implies that the first of these responses – reducing debt/interactions – is stabilizing. The second, however – reducing the strength of local dynamics – is destabilizing.

Conclusion

This paper presented a simple mathematical criterion for the stability of dynamical systems as interactions are added between subsystems. If the number and strength of interactions between subsystems grows too large, the criterion identifies a threshold of connectivity beyond which the dynamics are necessarily unstable. This result extends the

7

May theorem for random networks to structured complex networks. A number of artificial and naturally occurring complex systems are subject to this threshold. Indeed, any system within which the number and strength of interactions increase over time, without an attending increase in the strength of local stabilizing dynamics, will inevitably approach the interaction instability threshold. An interaction-driven core-halo instability lies at the heart of both financial and gravitational collapse, and may well underlie the current European sovereign debt crisis. Once the instability threshold is passed, equilibrium becomes unstable, and the dynamics becomes nonlinear [13-14] and catastrophic [15]. Rapidly growing fluctuations overwhelm physical and financial stabilization mechanisms. Stars collapse, and investments made according to the Black-Scholes equation [10] result in financial black holes.

References

[1] D.G. Luenberger, *Introduction to Dynamic Systems: Theory, Models and Applications,* Wiley, New York, (1979).

[2] J.-J.E. Slotine, W. Li, *Applied Nonlinear Control,* Prentice Hall, Englewood Cliffs (1991).

[3] S. Strogatz, *Nonlinear Dynamics and Chaos,* Perseus Books, Cambridge (1994).

[4] R. May, *Nature* **238**, 413 (1972).

[5] M. Newman, D. Watts, A.-L. Barabási, *The Structure and Dynamics of Networks,* (Princeton University Press, 2006).

[6] D. Lynden-Bell, R. Wood, *Mon. Not. R. Astr. Soc.* **138**, 495-525 (1968).

[7] K. Soramäki, M. L. Bech, J. Arnold, R.J. Glass, W.E. Beyeler, *Physica A* **379**, 317-333 (2007).

[8] A.G. Haldane, R.M. May, *Nature* **469**, 351-355 (2011).

[9] M.E.J. Newman, *SIAM Rev.* **45**, 167-256 (2003).

[10] F. Black, M. Scholes, *J. Pol. Econ.* **81**, 637-654 (1973).

[11] J. D. Farmer, J. Geanakoplos, 'Power laws in economics and elsewhere,' Sante Fe Institute working paper, 2008:
    http://www.elautomataeconomico.com.ar/download/papers/Farmer-powerlaw3.pdf.

[12] A. Clauset, C.R. Shalizi, M.E.J. Newman, *SIAM Review 51 (4),* 661703 (2009); arXiv:0706.1062.

[13] D. Sornette, *Phys. Rep.* **378**, 1-98 (2003).

[14] W. Lohmiller, J.-J.E. Slotine, *Int. J. Control* **78**, 678-688 (2005).

[15] V.I. Arnold, *Catastrophe Theory,* 3rd edition, Springer-Verlag, Berlin (1992).

Supplementary material

Definitions of stability:

The definition of stability adopted here – all perturbations shrink – is stronger than stability in the sense of Lyapunov, in which perturbations can initially grow, but must eventually shrink [1-2]. In a time-independent linearized system, the latter requirement is equivalent to demanding that the eigenvalues of ∇g have negative real parts. If the system is stable in the sense of Lyapunov, but G has positive eigenvalues, then some perturbations will grow before eventually dying out. The primary reason for adopting the stronger definition of stability is that for a large, complex system, the magnitude of the positive eigenvalues of G will typically scale as the size of the system, so that non-infinitesimal perturbations grow large very rapidly, violating the linearity assumptions, and taking the nonlinear system to a new and unpredictable regime. This scaling of the eigenvalues follows directly from theorem 1. Consider a family of systems $G(n)$ of increasing system size n. If the ratio $\mathrm{tr}C^\dagger C / \sqrt{\mathrm{tr}A^2}\sqrt{\mathrm{tr}B^2}$ is equal to $1 + \epsilon$ for all n, and if the number of entries in A, B, C grow proportional to n then the size of the largest eigenvalue of $G(n)$ grows in proportion to system size. A second technical reason for adopting the stronger definition of stability is that ∇g and G typically vary with time along the trajectory whose stability is being evaluated. For such time dependent systems, the condition that the eigenvalues of ∇g have negative real parts at all times no longer guarantees stability, while the condition that G have negative eigenvalues continues to guarantee stability [2]. Accordingly, the stronger definition of stability that G is negative definite and that all perturbations shrink is an appropriate definition for the types of complex systems considered here.

Proof of theorem 1:

Let

$$G_{off} = \begin{pmatrix} 0 & C \\ C^\dagger & 0 \end{pmatrix}. \tag{S1}$$

The singular value decomposition for C implies that the eigenvalues of G_{off} are either zero, or come in pairs $\pm\lambda_j$, where the $\lambda_j \geq 0$ are the singular values of the matrix C. The $\pm\lambda_j$ eigenvectors of G_{off} take the form $\vec{g}_j^\pm = \begin{pmatrix} \vec{u}_j \\ \pm\vec{v}_j \end{pmatrix}$, where $\vec{u}_j$ and $\vec{v}_j$ are the left-singular and right-singular vectors for λ_j: $C\vec{v}_j = \lambda_j\vec{u}_j$, $C^\dagger\vec{u}_j = \lambda_j\vec{v}_j$.

Now look at vectors of the form $\vec{w} = \begin{pmatrix} \alpha\vec{u}_j \\ \beta\vec{v}_j \end{pmatrix}$, where α, β are real and non-negative.

10

Maximizing $\vec{w}^\dagger G \vec{w}$ over α, β yields $\vec{w}^\dagger G \vec{w} > 0$ when $\lambda_j^2 > a_j b_j$, where $a_j = \vec{u}_j^\dagger A \vec{u}_j$ and $b_j = \vec{v}_j^\dagger B \vec{v}_j$. So the system is unstable if $\lambda_j^2 > a_j b_j$ for any j. In particular, if $\sum_j a_j b_j < \sum_j \lambda_j^2$, then the system is unstable.

Note that a_j, b_j are the diagonal elements of A, B in the bases $\{\vec{u}_j\}$ for A's n_A-dimensional Hilbert space, and $\{\vec{v}_j\}$ for B's n_B-dimensional Hilbert space. Let $\vec{\mu}$ be the vector of eigenvalues of A, and $\vec{\nu}$ be the vector of eigenvalues of B. It is straigtforward to verify that the vectors $\vec{a}$ with components a_j and $\vec{b}$ with components b_j are related to $\vec{\mu}$ and $\vec{B}$ by doubly stochastic transformations: $\vec{a} = W_A \vec{\mu}, \vec{b} = W_B \vec{\nu}$. Convexity then implies that $|\vec{a}|^2 = \sum_j a_j^2 \leq |\vec{\mu}|^2 = \sum_j \mu_j^2 = \operatorname{tr} A^\dagger A$. Similarly, $|\vec{b}|^2 = \sum_j b_j^2 \leq |\vec{\nu}|^2 = \sum_j \nu_j^2 = \operatorname{tr} B^\dagger B$. These inequalities, combined with the Cauchy-Schwartz inequality, show that

$$\operatorname{tr} C^\dagger C > \sqrt{\operatorname{tr} A^\dagger A} \sqrt{\operatorname{tr} B^\dagger B}$$
$$\longrightarrow \sum_j \lambda_j^2 > |\vec{\mu}||\vec{\nu}| \geq |\vec{a}||\vec{b}| \geq \vec{a} \cdot \vec{b} = \sum_j a_j b_j, \tag{S2}$$

and the system is unstable. This proves the theorem.

Theorem 1 immediately implies a set of stability tests. Define $a^2 = (1/n_A^2)\operatorname{tr} A^2 = (1/n_A)\sqrt{\sum_{ij=1}^{n_A} |a_{ij}|^2}$ to be the average magnitude squared of the entries of A. a can be thought of as the 'strength' of A's stabilizing dynamics. Similarly, $b = (1/n_B)\sqrt{\operatorname{tr} B^2}$ gives the 'strength' of B's stabilizing dynamics, and $c = (1/\sqrt{n_A n_B})\sqrt{\operatorname{tr} C^\dagger C}$ is the strength of the potentially destabilizing dynamics of the interactions. Theorem 1 is equivalent to the statement that when $c^2 > ab$, interactions cause instability: the system is unstable if the strength of the destabilizing interaction dynamics is greater than geometric mean of the strengths of the stabilizing local dynamics. Similarly, $c > (a + b)/2 \geq \sqrt{ab}$, and $c^2 > a^2/2 + b^2/2 \geq ab$ also imply instability.

The bound of theorem 1 is tight in the sense that for a given C, and fixed a, b such that $ab > c^2$, the entries of A and B can always be chosen to make the system stable. The 'minimal' strategy for attaining stability is align the eigenvectors of A and B with the left and right singular vectors of C. Arrange $\vec{\mu} \propto \vec{\nu}$, and $\mu_j \nu_j = \lambda_j^2 + \epsilon$. Then the system is stable and

$$\sqrt{\operatorname{tr} A^\dagger A} \sqrt{\operatorname{tr} B^\dagger B} = \operatorname{tr} C^\dagger C + O(\epsilon). \tag{S3}$$

Linearized equations for the gravo-thermal catastrophe:

Consider a system bound together by gravitation such as a star, nebula, or galaxy. Such systems typically possess a core-halo structure [6]. Let E_A, T_A be the energy and temperature of the core, and C_A its specific heat. Similarly, let E_B, T_B and C_B be the energy, temperature and specific heat of the halo. $\alpha \geq 0$ gives the linearized rate of energy transfer between core and halo as a function of their temperature difference. $\eta \geq 0$ governs heat production in the core, due, e.g., to nuclear reactions, and $\zeta \geq 0$ governs heat loss from the halo to space beyond. The linearized equations of motion are

$$\frac{dE_A}{dt} = C_A \frac{dT_A}{dt} = \alpha(T_B - T_A) + \eta T_A, \quad \frac{dE_B}{dt} = C_B \frac{dT_B}{dt} = \alpha(T_A - T_B) - \zeta T_B. \quad (S4)$$

Eliminating E_A, E_B then yields equation (4) of the text.

Instability of the bank transfer network:

To derive the interaction-driven instability threshold for the bank transfer network, divide the network into the set of banks with above-average numbers of links and volumes (the core, system $\mathcal{A}$), and those with below average numbers of links and volumes (the halo, system $\mathcal{B}$). Let v_{ij} be the measured volume flow from bank j to bank i during a particular reporting period. To construct a linearized dynamics would require knowledge of V_j, the amount of funds held in bank j, together with the rates v_{jj} of creation and consumption of funds within the j'th bank. These numbers are not available from the data set analyzed in [7]. Even though V_j and v_{jj} are unknown, however, the stability analysis of theorem 1 can still be applied. In the linearized dynamics, the matrix ∇g has entries v_{ij}/V_j, and the Hermitian gradient G has entries $g_{ij} = (v_{ij}/V_j + v_{ji}/V_i)/2$. The matrices A, B, and C, are derived from G as before: A governs the Hermitian dynamics within the core, B governs the dynamics within the halo, and C governs flows between core and halo. Let κ_A be the fraction of non-zero terms within A, so that $\kappa_A n_A^2$ is the number of links within the core. Define κ_B and κ_C in the same way. Let $\tilde{a}^2$ be the average magnitude squared of a non-zero term in A: $\tilde{a}^2 = a^2/\kappa_A$, where as above $a = (1/n_A)\sqrt{\mathrm{tr}A^\dagger A}$ is the average strength of all the terms in A, including those that are zero. Similarly, $\tilde{b}^2 = b^2/\kappa_B$, $\tilde{c}^2 = c^2/\kappa_C$ are the average magnitude squared of non-zero terms in B and C. Theorem 1 then implies that the dynamics are unstable if

$$\kappa_C \tilde{c}^2 > \sqrt{\kappa_A \kappa_B} \tilde{a}\tilde{b}. \quad (S5)$$

The disassortative nature of the network [9] now sets the stage for connectivity-driven instability: when the strength of internal stabilizing dynamics is insufficient, the dynamics

of the network gives rise to unstable and increasing flows between core and halo. The disassortative nature of the network implies that there are fewer internal links within the core, $\kappa_A n_A^2$, and within the halo, $\kappa_B n_B^2$, than there are between core and halo, $\kappa_C n_A n_B$. That is, disassortativity implies $\kappa_C^2 > \kappa_A \kappa_B$. Equation (S5) then implies that if all links correspond to terms of approximately the same size, the system is unstable. Stability requires that the strength of links within the core be significantly higher than the strength of links between core and halo, a feature observed in the actual network (3/4 of the volume of transfers occurs within the core).

If the system is marginally stable, so that $\kappa_C \tilde{c}^2 \approx \sqrt{\kappa_A \kappa_B} \tilde{a} \tilde{b}$, then any dip in the connectivity or the strength of connections within the core will drive it unstable.
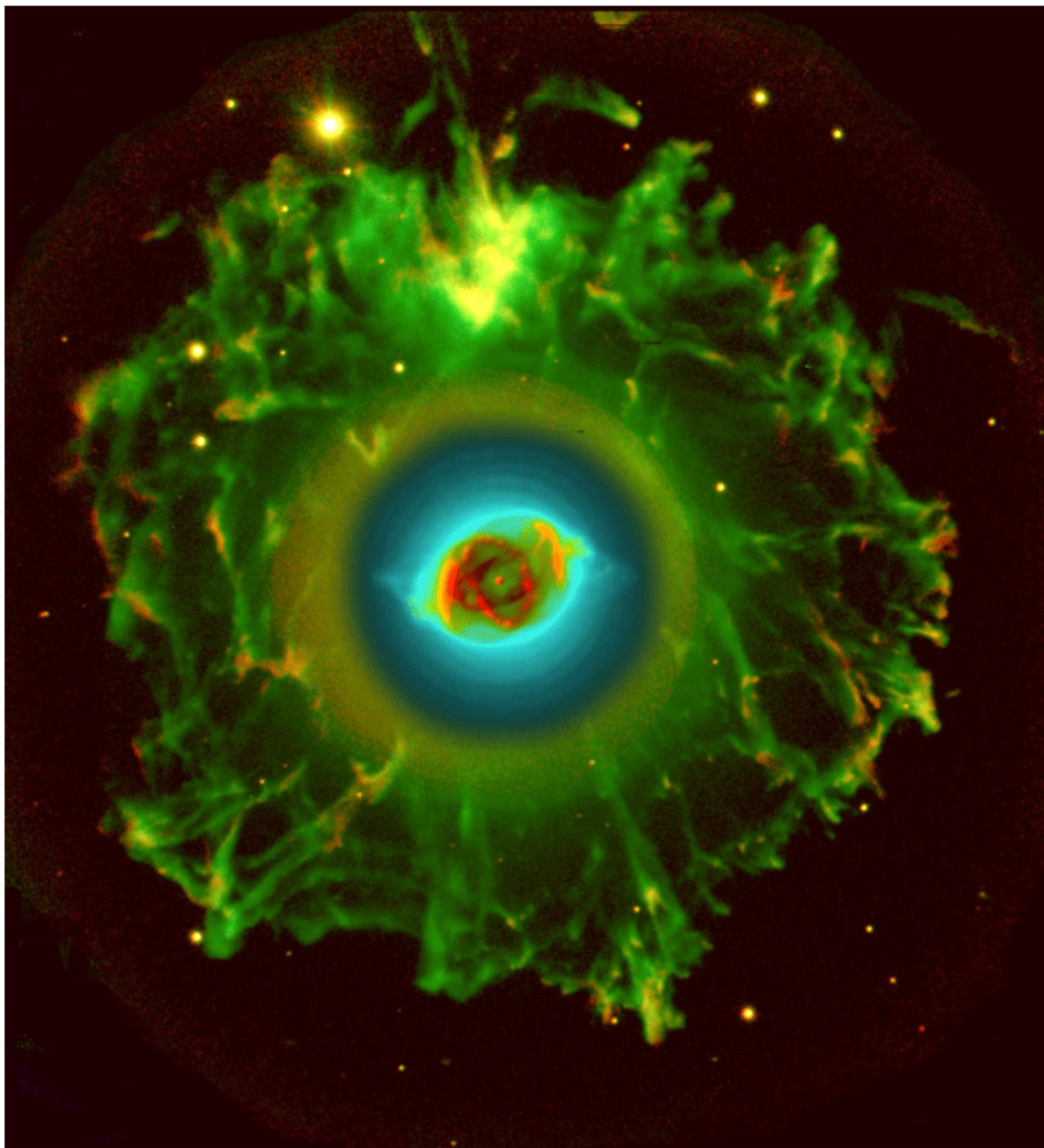
13

Figures

Figure 1: After. Cat's Eye nebula – the aftermath of a gravo-thermal catastrophe
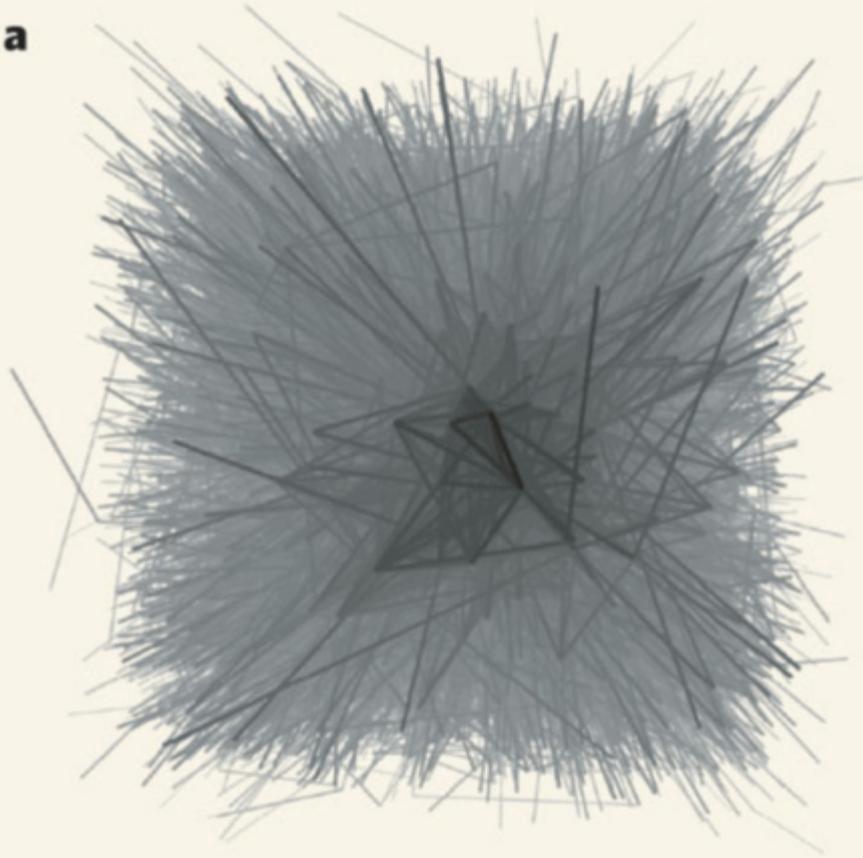
When a star burns through its nuclear fuel, the outgoing radiation pressure no longer suffices to support the star's weight, and it collapses. This collapse is an example of an interaction-driven core-halo instability, characterized by unstable flows of energy between core and halo, clearly seen in this image by Romano Corradi from the Nordic Optical Telescope.

Figure 2: Before. The interbank transfer network in 2007

The US interbank transfer network consists of over 6600 financial institutions making over 70,000 daily transfers. As mapped here from pre-financial crisis data, the network exhibits a pronounced core-halo structure (2a: reprinted with permission from reference [7]). The core of 66 institutions (2b) includes an inner core of 25 fully connected institutions, and accounts for 3/4 of the transfer volume. As with the gravo-thermal catastrophe, a slowdown of the core can drive the entire system unstable, leading to the financial analogue of gravitational collapse ('the black hole of finance').

14

EFTA00607109