Proceedings of the 12th European Conference on Combinatorics, Graph Theory and Applications
EUROCOMB'23
Prague, August 28 - September 1, 2023

Dispersion on the Complete Graph

(Extended abstract)

Umberto De Ambroggio* Tamás Makai* Konstantinos Panagiotou*

Abstract

We consider a synchronous process of particles moving on the vertices of a graph G, introduced by Cooper, McDowell, Radzik, Rivera and Shiraga (2018). Initially, M particles are placed on one vertex of G. At the beginning of each time step, for every vertex inhabited by at least two particles, each of these particles moves independently to a neighbour chosen uniformly at random. The process ends at the first step when no vertex is inhabited by more than one particle.

Cooper et al. showed that when the underlying graph is the complete graph on n vertices, then there is a phase transition when the number of particles $M = n/2$. They showed that if $M < (1-\varepsilon)n/2$ for some fixed $\varepsilon > 0$, then the process finishes in a logarithmic number of steps, while if $M > (1+\varepsilon)n/2$, an exponential number of steps are required with high probability. In this paper we provide a thorough analysis of the distribution of the dispersion time in the barely critical regime, where $\varepsilon = o(1)$, and describe the fine details of the transition between logarithmic and exponential time. As a consequence of our results we establish, for example, that the dispersion time is in probability and in expectation $\Theta(n^{1/2})$ when $|\varepsilon| = O(n^{-1/2})$, and provide qualitative bounds for its tail behavior.

DOI: https://doi.org/10.5817/CZ.MUNI.EUROCOMB23-046

1 Introduction

We consider the synchronous *dispersion process* introduced by Cooper, McDowell, Radzik, Rivera and Shiraga [1]. The process evolves in discrete time. It involves particles that move between vertices of a given graph G. A particle is called *happy*, if there are no other

*University of Munich, Department of Mathematics, Theresienstr. 39, 80333 Munich, Germany. E-mail: {deambrog,makai,kpanagio}@math.lmu.de. Supported by ERC Grant Agreement 772606-PTRCSP.

particles on the same vertex, otherwise it is *unhappy*. Initially, M particles are placed on some vertex of G. In every (discrete) time step, all unhappy particles move simultaneously and independently to a neighbouring vertex selected uniformly at random. Happy particles do not move. The process terminates at the first time at which all particles are happy. This (random) time is denoted by $T_{G,M}$ and it is called the *dispersion time*; it constitutes the main object of interest here.

In [1] the authors studied this process on several graphs, and established results concerning $T_{G,M}$ and the *dispersion distance*, which is the maximum distance of any particle from the origin at dispersion (that is, at step $T_{G,M}$). One of the main focus in [1] is the behaviour when the underlying graph is the complete graph with loops, which we will denote by K_n. The most general results come from considering a *lazy* variant of the dispersion process, which was shown to disperse more particles in a smaller number of steps. More precisely, in this lazy version any unhappy particle moves with probability $q \in (0, 1]$ and stays at its current location with probability $1 - q$.

The main result of [1] regarding $T_{K_n,M}$ is that there are constants $c, C > 0$ such that if $M = (1 - q/2 - \alpha)n$ for any $\alpha > 0$ that may depend on n, then

$$T_{K_n,M} \leq C(q\alpha)^{-1} \log(n) \text{ with probability at least } 1 - O(1/n), \tag{1}$$

whereas when $M = n(1 - q/2 + \alpha)$, then

$$T_{K_n,M} \geq e^{cnq^2\alpha^3} \text{ with probability at least } 1 - e^{-cnq^2\alpha^3}. \tag{2}$$

The above statements leave several questions open. Indeed, corresponding bounds for the lower and upper tails of $T_{K_n,M}$ were not provided. It is not clear, moreover, what the actual behavior is when M is close to $n/2$, that is, when $M = (1 + \varepsilon)n/2$ for some $|\varepsilon| = o(1)$ and how the transition from logarithmic to exponential time quantitatively looks like. For example, (2) is not informative when $q = 1$ and $\alpha = o(n^{-1/3})$, as it essentially only states that the number of steps is at least one.

Since we deal exclusively with the complete graph, in the following we will write $T_{n,M} = T_{K_n,M}$. Our main contribution is a thorough and precise analysis of the dispersion process when, as above, we assume that $M = (1 + \varepsilon)n/2$ and $|\varepsilon| = o(1)$. Then we establish that the process exhibits three qualitatively different behaviours based on the asymptotics of ε, where, informally speaking, $T_{n,M}$ smoothly changes from $|\varepsilon|^{-1} \log(\varepsilon^2 n)$ to $n^{1/2}$ and then to $\varepsilon^{-1} e^{\Theta(\varepsilon^2 n)}$; in particular, $T_{n,M} = \Theta(n^{1/2})$ for $M = n/2$. We begin with providing the upper bounds on the distribution of $T_{n,M}$.

Theorem 1.1. *There is a $C > 0$ such that the following is true for sufficiently large n and all $A \geq 1$. Let $\varepsilon = o(1)$ and $M = (1 + \varepsilon)n/2$. If $\varepsilon < -en^{-1/2}$, then*

$$\mathbb{P}\big(T_{n,M} > AC|\varepsilon|^{-1} \log(\varepsilon^2 n)\big) \leq e^{-(A-1)}.$$

Moreover, if $|\varepsilon| \leq en^{-1/2}$, then

$$\mathbb{P}\big(T_{n,M} > ACn^{1/2}\big) \leq e^{-(A-1)}.$$

Finally, if $\varepsilon > en^{-1/2}$*, then*

$$\mathbb{P}\big(T_{n,M} > A\varepsilon^{-1}e^{C\varepsilon^2 n}\big) \le e^{-(A-1)}.$$

The next main result establishes lower bounds for these ranges as well. When $\varepsilon \le en^{-1/2}$ these match, and in the last case we show that the exponential term is of the same order.

Theorem 1.2. *There is a* $c > 0$ *such that the following is true for sufficiently large* n *and all* $A \ge 1$*. Let* $|\varepsilon| = o(1)$ *and* $M = (1+\varepsilon)n/2$*. If* $\varepsilon < -en^{-1/2}$*, then*

$$\mathbb{P}\big(T_{n,M} \le c|\varepsilon|^{-1}\log(\varepsilon^2 n)/A\big) \le A^{-1}.$$

Moreover, if $|\varepsilon| \le en^{-1/2}$*, then*

$$\mathbb{P}\big(T_{n,M} \le cn^{1/2}/A\big) \le A^{-1}.$$

Finally, if $\varepsilon > en^{-1/2}$*, then*

$$\mathbb{P}\big(T_{n,M} \le \max\{e^{c\varepsilon^2 n}, c\varepsilon^{-1}/A\}\big) \le \min\{e^{-c\varepsilon^2 n}, A^{-1}\}.$$

Let us discuss briefly some consequences of our results. First of all, the two theorems combined imply that in probability

$$T_{n,M} = \Theta(|\varepsilon|^{-1}\log(\varepsilon^2 n)) \quad \text{if} \quad \varepsilon < -en^{-1/2},$$

and

$$T_{n,M} = \Theta(n^{1/2}) \quad \text{if} \quad |\varepsilon| = O(n^{-1/2}).$$

In particular, when $M = n/2$ we obtain that $T_{n,M} = \Theta(n^{1/2})$ in probability. For larger ε, we obtain the slightly weaker uniform estimate that in probability

$$\log(T_{n,M}) = \Theta(\varepsilon^2 n + \log n) \quad \text{if} \quad \varepsilon = \omega(n^{-1/2}).$$

This estimate can be improved as soon as $e^{c\varepsilon^2 n} \ge \varepsilon^{-1}$, that is, when $\varepsilon = \Omega((\log n/n)^{1/2})$; after this point the maximum in Theorem 1.2 will be $e^{c\varepsilon^2 n}$ and so, in fact, for such ε we obtain that even $\log(T_{n,M}) = \Theta(\varepsilon^2 n)$ in probability.

Apart from these estimates we can also use our main theorems to obtain information about, for example, the expectation of $T_{n,M}$. In particular, Theorem 1.1 guarantees that $T_{n,M}$ has an exponential(-ly thin) upper tail and so $T_{n,M}$ is integrable; we readily obtain that

$$\mathbb{E}[T_{n,M}] = \Theta(|\varepsilon|^{-1}\log(\varepsilon^2 n)) \text{ if } \varepsilon \le -en^{-1/2}, \qquad \mathbb{E}[T_{n,M}] = \Theta(n^{1/2}) \text{ if } |\varepsilon| = O(n^{-1/2}),$$

and

$$\log \mathbb{E}[T_{n,M}] = \Theta(\varepsilon^2 n + \log n) \text{ if } \varepsilon = \omega(n^{-1/2}).$$

Further Related work The dispersion process was also studied by Frieze and Pegden [3], who sharpened the result on the dispersion distance on L_∞, which denotes the infinite line. In particular, it was shown in [1] that with high probability, the dispersion distance on L_∞ for n particles is $O(n \log n)$; in [3] the logarithmic factor was eliminated. A similar setup was considered by Shang [8], where the author studied the dispersion distance in a non-uniform dispersion process in which an unhappy particle moves at the next time step to the right with probability p_n and to the left with probability $1 - p_n$, independently of other particles.

Processes where particles move on the vertices of a graph have been widely studied over the past decades; we refer the reader to [1] for references. Concerning processes whose scope is to *disperse* particles on a discrete structure, arguably the best known such model is the Internal Diffusion Limited Aggregation (IDLA, for short); see [2] and [4]. In this model, particles sequentially start (one at a time) from a specific vertex designated as the origin. Each particle moves randomly until it finds an unoccupied vertex; then it occupies it forever (meaning that it does not move at subsequent process steps). When a particle stops, the next particle starts moving. We emphasize that whenever a particle jumps to an occupied vertex, it just keeps moving without activating the occupant particle. In the dispersion process, on the other hand, when a (happy) particle standing alone on a node is reached by another particle, it is reactivated and keeps moving until it becomes happy again.

2 Proof Ideas

In the proof we begin with studying the expected change in the number of unhappy particles in every step. Let us write H_t and U_t for the number of happy and unhappy particles at the beginning of step t; in particular, $U_0 = M$ and $H_0 = 0$ and $U_t + H_t = M$ for all $t \in \mathbb{N}_0$. Then it turns out that

$$\mathbb{E}\left[U_{t+1} | U_t\right] = H_t \left(1 - \left(1 - \frac{1}{n}\right)^{U_t}\right) + U_t \left(1 - \frac{n - H_t}{n}\left(1 - \frac{1}{n}\right)^{U_t - 1}\right), \quad t \in \mathbb{N}_0. \tag{3}$$

The two summands correspond to the number of particles counted in H_t that become unhappy and to the number of particles counted in U_t that remain unhappy in step $t + 1$. Recall that we write $M = (1 + \varepsilon)n/2$ and assume that U_t is not too big, say $U_t \leq \delta n$ for some small $\delta > 0$. Then a quick calculation reveals that

$$\mathbb{E}\left[U_{t+1} | U_t\right] = (1 + \varepsilon)U_t - \Theta(U_t^2/n). \tag{4}$$

So, as long as U_t is (much) larger than $|\varepsilon|n$, then U_{t+1} will be (much) smaller than U_t in expectation. In other words, when there are 'many' unhappy particles, the expected number of unhappy particles in the next step decreases significantly. However, this is no longer the case when there are only a 'few' unhappy particles, that is, less than $O(|\varepsilon|n)$. In this case the number of unhappy particles is expected to either decrease only by a slight

amount, which in particular is problematic when the expected decrease $\mathbb{E}[U_{t+1} - U_t | U_t]$ is smaller than one, or when $\varepsilon > 0$, where we can expect that the number of unhappy particles even *increases*. We will use different methods to analyse the trajectory of U_t depending on the range of ε and whether we are considering an upper or a lower bound. In particular, when $|\varepsilon|$ is not too large we will see that we can compare the situation to a very slightly biased random walk, and so we will end up with an $n^{1/2}$ term, while in the other cases our walk will have a positive/negative drift and the exponential/logarithmic term will emerge.

2.1 Upper tail

Our approach for establishing Theorem 1.1 is to find a lower bound on the probability that when starting with an *arbitrary* number of unhappy particles, the process will stop within a certain number of steps. By splitting the time interval under consideration into a disjoint union of smaller intervals and using the Markov property of the process, we can apply the above bound repeatedly on these smaller sections to achieve an exponentially decreasing upper bound on the probability that the process is still not finished.

When there are many unhappy particles we use drift analysis to analyse the process; we refer the reader to [5] for an excellent introduction and description of the method. Roughly speaking, drift analysis provides an estimate for the expected duration of a homogeneous Markov-process over a discrete state space, when the expected value of the conditional one step change is known for every element in the state space. With (4) at hand we can apply the method to deduce bounds for the probability that dispersion leaves us with many unhappy particles after a certain number of steps.

Once there are only a few unhappy particles left we change our approach. After this point we bound from above the number of unhappy particle with another random process, which we call the *binomial process*. More specifically, beginning with some initial value B_0, we define a random process by setting $B_{t+1} = 2\mathrm{Bin}(B_t, M/n)$, $t \in \mathbb{N}_0$. The quantity B_t provides an upper bound for the number of unhappy particles after t steps, as the probability that an unhappy particle lands on the same vertex as any other particle is at most M/n, and in that case we account for two unhappy particles. As the number of unhappy particles is small, it is rare for two unhappy particles to land on the same vertex, making this coupling relatively tight.

The binomial process is equivalent to B_0 independent copies of a Galton-Watson branching process that have no offspring with probability $1 - M/n$ and two offspring with probability M/n. A simple inductive argument implies that the size of the k-th generation of these B_0 branching processes has the same distribution as B_k.

In the next step we estimate the probability that a single copy of the branching process survives for at least k generations; denote this probability by x_k. Then $x_0 = 1$ and moreover

$$x_{k+1} = \frac{M}{n}\left(2x_k - x_k^2\right),$$

as in order for the branching process to survive for $k+1$ generations, the root has to have 2 children, and at least one of these children has to survive for at least k generations.

Recall the well-known property concerning the survival probability of a Galton-Watson branching process, namely that it is 0 if the expected number of offspring is at most 1, and it is bounded away from 0 when the expected offspring is larger than 1 (see e.g. [7]). Clearly x_k tends to the survival probability as $k \to \infty$, and moreover, $\lim_{k\to\infty} x_k = o(1)$ as $n \to \infty$, as the expected number of children is ~ 1.

So far we have referred to *many* and *few* unhappy particles, without mentioning the level where the change occurs. The exact value is determined by carefully balancing several properties. For the branching process argument, we would like that the number of steps is such that x_k is sufficiently close to its limit. In addition we would like the number of steps that we study using the branching process to match the number of steps we analyse using drift analysis, so as to reach an optimal bound. This leads to two different regimes, namely

- roughly $n^{1/2}$ steps, when $|\varepsilon| \leq en^{-1/2}$;
- roughly $|\varepsilon|^{-1}$ steps, when $|\varepsilon| > en^{-1/2}$,

which coincide with the three regimes in the main theorem.

The branching process behaves similarly in all regimes, as the survival probability is $o(1)$. However, the rate of convergence in n becomes slower as ε increases and thus the probability that all independent copies of this branching process die out (in the required number of steps) goes from almost certain, when $\varepsilon < en^{-1/2}$ to a constant when $|\varepsilon| \leq en^{-1/2}$, to exponentially decreasing when $\varepsilon > en^{-1/2}$. Recall that we use these probabilities as the basis of a geometric distribution, which leads to the upper bounds in Theorem 1.1.

2.2 Lower tail

Now we consider the lower bound. In this case we first show that we can mostly ignore what happens when the number of unhappy particles is large, the only exception is that we have to ensure that it is unlikely that most of the unhappy particles become happy in any single step. In order to achieve this, note that U_{t+1} is a function of U_t and the vertices to which the unhappy particles counted in U_t jump to. Then we consider the Doob martingale induced by exposing the individual destinations (of the unhappy particles) one at a time and show concentration around its expectation with Azuma-Hoeffding, which yields the desired property that the number of unhappy particles does not decrease too quickly.

Having established this, we proceed similarly to the analysis for the upper tail. Let $B'_0 \in \mathbb{N}$ and consider the binomial process with different parameters defined by

$$B'_{t+1} = 2\mathrm{Bin}(B'_t, (M - 2K)/n), \quad t \in \mathbb{N},$$

where $K \in \mathbb{N}$ is arbitrary but fixed. Then, as long as $U_t \leq K$, we have that B'_t is a lower coupling for U_t, as there are at least $M - 2K$ happy particles at the end of the corresponding step. We can analyse the binomial process using branching processes as for the upper bound, providing a lower bound with the right order of magnitude when

$\varepsilon \leq en^{-1/2}$. However, when $\varepsilon > en^{-1/2}$, due to the $-2K$ term the associated branching process becomes subcritical, and hence we are only able to show that the process still runs after $\Omega(\varepsilon^{-1})$ steps, which provides the corresponding term in the max in Theorem 1.2.

In order to obtain the term involving the exponential in $\varepsilon^2 n$, we use an alternative approach, adapting the argument of Theorem 2.6 in Lengler and Steger [6]. Note that this term only appears when $\varepsilon > en^{-1/2}$, thereby whenever U_t is small, (4) indicates that in the following step U_t will increase in expectation. In such a case it is unlikely that U_t decreases, and consequently many steps are required before the prcess stops.

References

[1] C. Cooper, A. McDowell, T. Radzik, N. Rivera, and T. Shiraga. Dispersion processes. *Random Structures Algorithms*, 53(4):561–585, 2018.

[2] P. Diaconis and W. Fulton. A growth model, a game, an algebra, Lagrange inversion, and characteristic classes. *Rend. Sem. Mat. Univ. Politec. Torino*, 49(1):95–119, 1991.

[3] A. Frieze and W. Pegden. A note on dispersing particles on a line. *Random Structures Algorithms*, 53(4):586–591, 2018.

[4] G. F. Lawler, M. Bramson, and D. Griffeath. Internal diffusion limited aggregation. *Ann. Probab.*, 20(4):2117–2140, 1992.

[5] J. Lengler. Drift analysis. In B. Doerr and F. Neumann, editors, *Theory of Evolutionary Computation: Recent Developments in Discrete Optimization*, Nat. Comput. Ser., pages 89–131. Springer, 2020.

[6] J. Lengler and A. Steger. Drift analysis and evolutionary algorithms revisited. *Combin. Probab. Comput.*, 27(4):643–666, 2018.

[7] R. Lyons and Y. Peres. *Probability on trees and networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016.

[8] Y. Shang. Longest distance of a non-uniform dispersion process on the infinite line. *Inform. Process. Lett.*, 164:106008, 5, 2020.