Crowd Flux Theory:
A Statistical Approach to Crowd Dynamics

Taiyo Tagami

April 23, 2025

ABSTRACT: This paper covers the development and applications of crowd flux theory (CFT). CFT is a mathematical model of how individuals, or densities of individuals, fluctuate in space and over time. Developed off of the foundations of classical physics and density-functional fluctuation theory (DFFT) developed in Méndez-Valderrama et al. 2018 [4], this model potentially allows for efficient simulation and analysis of high and low crowd density situations, as well as the transition between.

Contents

1 Introduction **1**

2 Background **2**

3 The Crowd Flux Theory Model **3**

4 Improving the Efficiency of CFT **7**

5 Discussion **9**

6 Future Works **9**

1 Introduction

Over the past few years, the world has experienced a number of massive crowd disasters and it is highly plausible that they will continue to happen in the future unless preventive measures are taken. Furthermore, the COVID-19 pandemic has ignited the world's interest in crowd analysis as a measure of reducing infection rates. These factors make crowd dynamics more relevant than ever in the present day.

However, human behavior is highly complex; most methods today can only predict human behavior under specific conditions and with debatable certainty. Popular microscopic agent-based models—while they have the potential to be greatly nuanced and complex—are limited in their accuracy and cannot be scaled to large crowds due to computational limitations [7]. Macroscopic models, such as the fluid dynamics model, are computationally efficient but are not very precise in their predictions and will fall apart when individualistic behaviors are not negligible [8] [7]. Therefore, I took to developing a versatile and efficient model for studying and predicting crowd behavior, which I have named "crowd flux theory."

2 Background

Density-functional fluctuation theory

The CFT model works off of the foundation of the DFFT model, which assumes that individuals have a quantifiable dissatisfaction with being in certain locations and will move to different locations accordingly. The DFFT model assigns the function $V(\mathbf{x})$—the vexation function—as the intrinsic desirability of location $\mathbf{x}$, where $\mathbf{x}$ becomes more preferable as $V(\mathbf{x})$ decreases [4]. The function $f'(n)$ represents the desirability of a location with areal crowd density n [4]. The density of location $\mathbf{x}$ is represented by $n(\mathbf{x})$ such that the preferability of location $\mathbf{x}$ is given by $V(\mathbf{x}) + f'(n(\mathbf{x}))$ [4].

The probability of an individual going from location $\mathbf{x}$ to $\mathbf{x}'$ can be described as the probability function

$$\mathbb{P}(\mathbf{x}, \mathbf{x}') = \frac{1}{1 + \exp(\Delta H)} \tag{1}$$

where ΔH is the difference in the preferability of locations $\mathbf{x}$' and $\mathbf{x}$: $V(x') + f'(x') - (V(x) + f'(x))$ [4]. This is the most natural description of the probability function, where the probability an individual goes to a new location approaches 0 the less preferable the new location is in contrast to their current location, the probability an individual goes to a new location approaches 1 the more preferable the new location is in contrast to their current location, and the probability is $\frac{1}{2}$ when the two locations are equally preferable.

By extracting global behaviors by integrating ΔH over dn and dA, Mendez-Valderrama, et al. found that the probability that density $n(\mathbf{x})$ occurs is given by the probability density functional

$$P[n(\mathbf{x}] = Z^{-1} exp\left(-\int\int (V(\mathbf{x}) + f'(n(\mathbf{x})))dndA\right) \tag{2}$$

where Z is the normalization constant and $\int ndA$ is the number of individuals in a given location [4]. (For a more detailed overview of DFFT, see Méndez-Valderrama, et al. 2018 [4])

This is then discretized into the following form using discrete counts of individuals in equal-size bins, which gives the probability of N individuals being in bin b

$$P_b(N) = z_b^{-1}\frac{1}{N!}exp(-(Nv_b + f_N)) \tag{3}$$

where z_b is the bin-independent normalization constant, v_b is the average value of the vexation within the bin, and $f_N = Af(N/A)$.

The CFT model develops a very similar discretized model of the probability of N individuals being in bin b using additional parameters besides the frustration and vexation functions and a continuous form of crowd density.

Crowd-dynamic interactions

3 The Crowd Flux Theory Model

The momentum function

While the DFFT model is not focused on simulating individual movement but instead on crowd behavior, it provides a great foundation for analyzing individualistic movement. This can be done by modifying the bin-independent probabilistic function to consider two separate locations to allow for the simulation of movement.

This is done through the addition of the momentum dissatisfaction function $M(\mathbf{x}, \mathbf{x}', \mathbf{p})$ that represents how dissatisfying location $\mathbf{x}'$ is for an individual with momentum $\mathbf{p}(\mathbf{x})$ at position $\mathbf{x}$ when the two locations are equally preferable otherwise. The momentum function $\mathbf{p}(\mathbf{x})$ is derived from the definition of momentum in classical physics. In classical physics, the momentum of an object is $\mathbf{p} = m\mathbf{v}$, where m is the mass of the object and $\mathbf{v}$ is the velocity of the object. Here, I define the momentum of a crowd to be

$$\mathbf{p} = I\mathbf{v}\int ndA \tag{4}$$

where I is the resistance an individual has to altering movement [1], $\int ndA$ represents the number of individuals at the location, and $\mathbf{v}$ is the average velocity of all the individuals in the location. The velocity of an individual can be estimated to be $\frac{\mathbf{x}'-\mathbf{x}}{t}$, where $\mathbf{x}'$ is their current location, $\mathbf{x}$ is their previous location, and t is the time they took to travel from $\mathbf{x}$ to $\mathbf{x}'$. $I\int ndA$ is analogous to m in the momentum equation from classical physics. It is important to note that, since individuals are influenced by the motion of the crowd in close proximity due to physical factors such as collision [8], I have assumed in this equation that all individuals in the same location will share the same momentum.

If there is only one individual being observed, $\int ndA = 1$ and $\mathbf{v}$ is the velocity of the individual. This allows for the crowd momentum equation to collapse into an individualistic momentum equation

$$\mathbf{p}_1 = I\mathbf{v} \tag{5}$$

that describes how, for individuals with higher inertia, a larger impulse—defined in classical physics to be the change in momentum and also the amount of force applied times the time it is applied over—is needed to change the velocity of the individual by a certain amount in comparison to an individual with lower inertia. By letting the difference of the sum of the density-based frustration and intrinsic vexation of locations function as the only force that is applied to an individual, an individual will resist changing their velocity to go to a more preferable location based on their inertia.

However, to apply momentum to the probabilistic foundations provided by DFFT, the momentum of individuals must be expressed in terms of the individual's dissatisfaction. Thus, I let the desirability of a location **x'** for an individual at location $\mathbf{x}$ with momentum $\mathbf{p}(\mathbf{x})$ and density $n(\mathbf{x})$ be defined by the momentum-dissatisfaction function $M[\mathbf{x}', \mathbf{x}, \mathbf{p}(\mathbf{x}), n(\mathbf{x})]$. The addition of this function to the difference in preferability such that

$$\Delta H = V(x') + f'(x') + M[\mathbf{x}', \mathbf{x}, \mathbf{p}(\mathbf{x}), n(\mathbf{x})] - (V(x) + f'(x) + M[\mathbf{x}, \mathbf{x}, \mathbf{p}(\mathbf{x}), n(\mathbf{x})]) \quad (6)$$

causes the probability function

$$\mathbb{P}(\mathbf{x}, \mathbf{x}') = \frac{1}{1 + \exp(\Delta H)}$$

(Eq. (1)) vary dependent on the individual being observed. This inserts individualism into DFFT, as well as a time component that allows CFT to track the movement of crowds over time.

By integrating ΔH over dn and dA, I let the net frustration effect $f(n(\mathbf{x})) = \int f'(n(\mathbf{x}))dn$ and net momentum-dissatisfaction $\mu[\mathbf{x'}, \mathbf{x}, \mathbf{p}(\mathbf{x}), n(\mathbf{x})] = \int M[\mathbf{x'}, \mathbf{x}, \mathbf{p}(\mathbf{x}), n(\mathbf{x})]dn$ and derived the probability density functional of location $\mathbf{x}'$ with respect to location $\mathbf{x}$ can be predicted as is done in Méndez-Valderrama, et al.'s work [4]

$$P[n(\mathbf{x'}), n(\mathbf{x})] = Z^{-1}\exp\left(-\int (V(\mathbf{x'})n(\mathbf{x}) + f(n(\mathbf{x'})) + \mu[\mathbf{x'}, \mathbf{x}, \mathbf{p}(\mathbf{x}), n(\mathbf{x})])dA\right) \quad (7)$$

12pt This allows for the probability of a certain density occurring at a location to be measured with respect to another location, implying the movement of individuals.

Momentum-based Crowd Interactions

One apparent flaw of Eq. (7) is that it does not consider any interactions between individuals not occupying exact same location, which in itself is impossible. Furthermore, the density function $n(\mathbf{x})$ cannot be calculated exactly without discretization in experimental settings. As done in Méndez-Valderrama's paper, the density function $n(\mathbf{x})$ can be discretized to equal the discrete counts of individuals over equal area bins [4]. The size of the bins must be large enough to capture interactions between individuals but small enough that the vexation within the bin does not vary significantly [4]. Eq. (7) can be discretized as

$$P_{b,c}(N) = z_{b,c}^{-1}\frac{1}{N!}exp(-(Nv_b + f_N + \mu_N)) \tag{8}$$

where

- N is the number of individuals in bin b
- v_b is the average vexation over bin b
- $f_N = Af(N/A)$ for area A
- $\mu_N = A\mu[b(\mathbf{x}), c(\mathbf{x}), \mathbf{p}_c, N/A]$ for area A
- locations of bin b and bin c as $b(\mathbf{x})$ and $c(\mathbf{x})$ respectively
- average momentum $\mathbf{p}_c$ in bin c

Similarly to the discretized probability function of DFFT, $z_{b,c}$ is a normalization constant for bin b with respect to bin c and $N!$ accounts for equivalent configurations among bins [4].

Grouped Behavior

In many cases, crowds do not behave homogeneously [7]. While some individual differences can become negligible in a large crowd, there are situations when the perceived vexation values for different locations can be very different for different groups. A good example of when such a scenario must be tested is when simulating a two-directional flow in a hallway [6]: one group of people might be walking down the hallway to their destination, while another group may be walking up the hallway to a different destination. To model such a situation, the perceived vexation values of locations must be different for the two groups. The vexation of a location closer to an individual's destination is generally lower than that of a location that is further, as locations closer to the destination tend to be preferable to locations further from the destination. A similar situation can hypothetically arise with the frustration and momentum-dissatisfaction values. To allow CFT to deal with this issue, the population being observed can be divided into m behavioral groups. Accordingly, I modified Eq. (6) such that

$$\Delta H = \sum_{i=1}^{m}(V_i(x') + f_i'(x') + M_i[\mathbf{x}', \mathbf{x}, \mathbf{p}(\mathbf{x}), n(\mathbf{x})] - (V_i(x) + f_i'(x) + M_i[\mathbf{x}, \mathbf{x}, \mathbf{p}(\mathbf{x}), n(\mathbf{x})])) \tag{9}$$

A probability density functional can then be derived , which can then be discretized into the form

$$P_{b,c}(N) = z_{b,c}^{-1}\frac{1}{N!}\exp\left(-\sum_{i=1}^{m}(N_i v_{b,i} + f_N, i + \mu_N, i)\right) \tag{10}$$

where

- N is the number of individuals in bin b
- $z_{b,c}$ is the normalization constant
- N_i is the number of individuals in group i in bin b such that $\sum_{i=1}^{m} N_i = N$
- $v_{b,i}$ is the group-dependent vexation of bin b for group i
- $f_{N,i} = Af_i(N_i/A)$ for area A, where f_i is the net frustration effect for individuals in group i
- $\mu_{N,i} = A\mu_i[b(\mathbf{x}), c(\mathbf{x}), \mathbf{p}_c, N_i/A]$ for area A, where μ_i is the net frustration effect for individuals in group i
- locations of bin b and bin c as $b(\mathbf{x})$ and $c(\mathbf{x})$ respectively
- average momentum $\mathbf{p}_c$ in bin c

This form of the discrete probability density functional is capable of capturing varying individual and group behaviors through the progression of time, allowing CFT to mimic microscopic agent-based simulations[1].

Application to Simulations

As the discrete CFT probability density functional (Eq. (10)) depends on two bins and their properties, it is not bin-independent. This means that they cannot independently fluctuate when applied to simulations. Instead, the probability N individuals move from bin c to b_ω given individuals can move from c to any bin in set $A = \{b_1, b_2, \ldots, b_{|A|}\}$ is defined:

$$\mathbb{P}_{b_\omega,c}(N) = \frac{P_{b_\omega,c}(N)}{(\sum_{i=\omega}^{|A|}(P_{b_i,c}(N)))} \tag{11}$$

`Note that` c `can be an element of set` A

Applied to simulations, each bin containing one or more individuals can run iterate this equation for each bin in set A. Assuming that individuals have a maximum travel speed, the bins that individuals can travel to from their current bin within a given time frame (the framerate of the simulation) can be defined. This provides the complete framework by which basic CFT-based simulations will run.

[1] Generally, microscopic agent-based simulations simulate each individual as an independent "agent" that moves with its own rules over time. For example, a simulation of 100 people walking down a corridor will have 100 distinct agents, each following instructions to walk down the corridor and interact to some (or possibly no) extent with other agents as the simulation progresses through time. CFT mimics agent-based models as they are the most versatile and precise models for simulating crowds—however, the drawback is that they are very computationally costly. [7]

4 Improving the Efficiency of CFT

Estimating the Factorial

Eq. (10) presents a significant challenge in computational efficiency: it contains the factorial of the number of individuals. This means that, while the equation itself must only be iterated once per bin, the calculation of the factorial value requires iteration for each individual present in the bin. This means that the big O time complexity of the CFT model as presented in Eq. (10) is $O(P \times N)$ where P is the precision of the bins (defined as the number of bins per unit area) and N is the number of individuals. The big O time complexity shows that, as the number of individuals is increased, the number of operations that must be performed scales proportionally[2].

To seamlessly transition between microscopic and macroscopic modeling, the Big O time complexity of the CFT model with respect to the number of individuals must be improved. This was done by using Rammanujan's approximation of the factorial to approximate the factorial as:

$$N! \approx \sqrt{\pi}\left(\frac{N}{e}\right)^N \sqrt{8N^3 + 4N^2 + N + 1/30}$$

Applying this to the CFT model, Eq. (8) can be approximated as

$$P_{b,c}(N) \approx z_{b,c}^{-1} \frac{1}{\sqrt{2\pi N}\left(\frac{N}{e} + \frac{1}{12eN}\right)^N} \exp(-(Nv_b + f_N + \mu_N)) \tag{12}$$

with the error indicated by n μ_n in Figure 1.

$n!$	$n! - \sigma_n$	$n! - \gamma_n$	$\rho_n - n!$	$n! - \mu_n$
10	30104	239.18	0.3116	**2.5416**
100	7.7739×10^{154}	6.4479×10^{151}	8.8211×10^{146}	$\mathbf{6.4815 \times 10^{148}}$
250	1.0774×10^{489}	3.5847×10^{485}	7.8705×10^{479}	$\mathbf{1.4368 \times 10^{482}}$
500	2.0334×10^{1130}	3.3858×10^{1126}	1.8603×10^{1120}	$\mathbf{6.7786 \times 10^{1122}}$
2500	5.4295×10^{7406}	1.8095×10^{7402}	3.9801×10^{7394}	$\mathbf{7.2395 \times 10^{7397}}$

Figure 1: Accuracy of factorial approximations [5]

[2]The number of operations done per bin scales proportionally. As this is also the case in DFFT, I judged that there is no need to improve the time complexity with respect to the precision of the bins.

(The approximation developed in Mortici 2011 is compared to Stirling's approximation $n!\ \sigma_n$, Gosper's approximation $n!\ \gamma_n$, and Ramanujan's approximation $n!\ \rho_n$. Mortici's approximation has better performance than Gosper's but worse than Ramanujan's. Thus, Ramanujan's approximation was chosen over Mortici's approximation.)

This form of the probabilistic requires no iteration for the number of individuals. Thus, when this is applied to Eq. (10), the CFT model has a big O time complexity of $O(P)$ for calculating the probability of a number of individuals moving from one bin to another. This means that the number of operations theoretically remains constant no matter how many individuals are being observed. However, it should be considered that double precision only functions for 64-bit decimal numbers. Working with large numbers, rounding must be done to prevent numbers from growing beyond 64-bit precision. The error scales as shown below

Better yet, one can create beforehand a lookup table of factorial values when dealing with discrete counts of individuals, which presents no error. This both the most efficient and precise method of dealing with the factorial, but in the case that non-discrete numbers are used and the factorial is exchanged for $N\Gamma(N)$, the approximation above is useful.

Logistic Approximation of Cumulative Distribution

In order to apply these mathematically-derived probabilities in a computer simulation, there must be a way to distribute individuals to various nodes using the given probability. My method of accomplishing this is to utilize the cumulative distribution of the CFT probabilistic function shown in Eq. (10). This CFT probabilistic follows a modified Poisson form, where λ is $\sum_{i=1}^{m}(N_i v_{b,i} + f_N, i + \mu_N, i)$. This can be approximated into a normal distribution form $N(\mu = \lambda, \sigma = \sqrt{\lambda})$ [3]. This then can be used to calculate the cumulative normal distribution $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{u^2}{2}\right\} du$ approximation of the cumulative distribution of Eq. (10). The logistic approximation to the cumulative normal distribution developed in Bowling, et al. 2009 [2] can then be used to create a simple logistic function that closely approximated the cumulative distribution of Eq. (10) in the form of:

$$F(z) = \frac{1}{1 + \exp(-(0.0705z^3 + 1.5976z))} \tag{13}$$

where $z = \frac{N-\mu}{\sigma}$ [2] and N is the number of individuals.

This logistic approximation can be reorganized into the following cubic equation

$$0.07056z^3 + 1.5976z + ln\left(\frac{1}{F(z)} - 1\right) = 0 \tag{14}$$

which follows the form of a depressed cubic. Using Tartaglia's method, this equation can be solved for z in terms of $F(z)$. Thus, by generating a random number between 0 and 1 for $F(z)$, one can solve for a specific value of z that satisfies Eq. (14). N can then be solved from z using the μ and σ. N can then be rounded to return the discrete number of individuals moving from one bin to another.

Using this method, the number of individuals moving from one bin to another can be closely approximated without the need to iterate for every individual in a bin. Thus, this improves the big O complexity of CFT simulation to $O(P)$.

5 Discussion

The CFT model demonstrates a potential for significant increases in the computational efficiency of crowd simulations. The CFT model is capable of simulating microscopic behaviors in low-density conditions and macroscopic behaviors in high-density conditions with greater computational efficiency than standard agent-based models. Microscopic-based models, while very precise, have a big O time complexity of $O(N)$, where N is the number of agents. This means that the number of operations done scales proportionally with the number of agents. The CFT model, however, has a big O notation of $O(P)$ where P is the precision of the bins. This means that the number of operations does not scale with the number of agents, but with the assigned precision of the bins. This is advantageous as the precision of the bins can be changed freely depending on the precision of the simulation required. The accuracy of the number of agents, on the other hand, can rarely be sacrificed for the sake of improving computational efficiency. This potentially makes the CFT model more versatile than standard agent-based methods.

The CFT model also has great potential in the field of machine learning. Through binning methods, the frustration and momentum-dissatisfaction functions can be estimated using object-recognition AI to extract the probabilities that people move from one location to another from a database of crowd footage. The estimated functions can then be applied to predict the movement of crowds across space in real-time. The model may provide insight into the behavioral patterns of organisms, such as social behaviors from the frustration function, laziness or resistance to changing one's course of action from the momentum-dissatisfaction function, and environmental preferences from the vexation function. Finally, the CFT model can allow for affordable, accurate predictions for crowd disaster risks and disease spread, making social events safer and saving lives.

6 Future Works

Testing the Model

Before CFT can be confidently used, its accuracy must be tested using real data. The CFT model, as presented in this paper, is purely theoretical. While it is arguably general enough to accurately simulate crowd behavior given appropriate vexation, frustration, and momentum-dissatisfaction functions, the difficulty of attaining these functions has yet to be fully tested and considered. Furthermore, if these functions are highly complex, the big O time complexity can significantly worsen. Thus, it is crucial that the CFT model is tested against real data in the future.

Applications to Machine-Learning

A strength of the CFT model is in its possible applicability to machine learning. From probabilities extracted from real crowd data, the functions can be defined purely mathematically. However, the efficiency of this process has yet to be tested. The DFFT model provides a good baseline for machine-learning applicability; the CFT and DFFT models can be compared to each other in terms of their machine-learning efficiency. One possible comparison that can be made is between the two models' prediction accuracy after a given time of training. If the CFT model proves more efficient than the DFFT model for machine learning, the CFT model may find great use in extracting mathematical functions that approximate the behavior of crowds.

References

[1] Carlos Alós-Ferrer, Sabine Hügelschäfer, and Jiahui Li. “Inertia and Decision Making”. In: *Frontiers in Psychology* 7 (2016).

[2] Shannon R. Bowling et al. “A Logistic Approximation to the Cumulative Normal Distribution”. In: *Journal of Industrial Engineering and Management* (1 July 2009). DOI: 10.3926/jiem.2009.v2n1.p114-127.

[3] Ivo D. Dinov. *Normal Approximation to Poisson(λ) Distribution*. Accessed: 2023-06-25. Feb. 2019. URL: http://www.socr.ucla.edu/Applets.dir/NormalApprox2PoissonApplet.html.

[4] J. Felipe Méndez-Valderrama et al. “Density-functional fluctuation theory of crowds”. In: *Nature Communications* 9 (3538 Aug. 2018).

[5] Christinel Mortici. “A substantial improvement of the Stirling formula”. In: *Applied Mathematics Letters* 24 (8 Aug. 2011), pp. 1351–1354.

[6] Van Romero, William D. Stone, and Julie Dyke Ford. “COVID-19 Indoor Exposure Levels: An Analysis of Foot Traffic Scenarios within an Academic Building.” In: *Transportation Research Interdisciplinary Perspectives* 7 (2020). DOI: 10.1016/j.trip.2020.100185.

[7] R. A. Saeed, Diego Reforgiato Recupero, and Paolo Remagnino. *Simulating crowd behaviour combining both microscopic and macroscopic rules*. Jan. 2021.

[8] Wouter van Toll et al. “SPH Crowds: Agent-Based Crowd Simulation up to Extreme Densities Using Fluid Dynamics”. In: *Computers & Graphics* (Aug. 2021), pp. 306–21. DOI: 10.1016/j.cag.2021.06.005.