2023.7.26 SIGIR'23 @ Taipei

A Critical Reexamination of Intra-List Distance and Dispersion

Naoto Ohsaka & **Riku Togashi**

(CyberAgent, Inc.)

https://todo314.github.io/ https://riktor.github.io/

Diversified recommendation

Many beyond-accuracy objectives
[Castells-Hurley-Vargas. 2015] [Kaminskas-Bridge. 2017] [Zangerle-Bauer. 2022]

- diversity, novelty, serendipity, coverage, ...
- Motivation: Risk of recommending over-specialized items to a user

We will focus on **diversity**
:= Internal differences btw. items recommended to a user

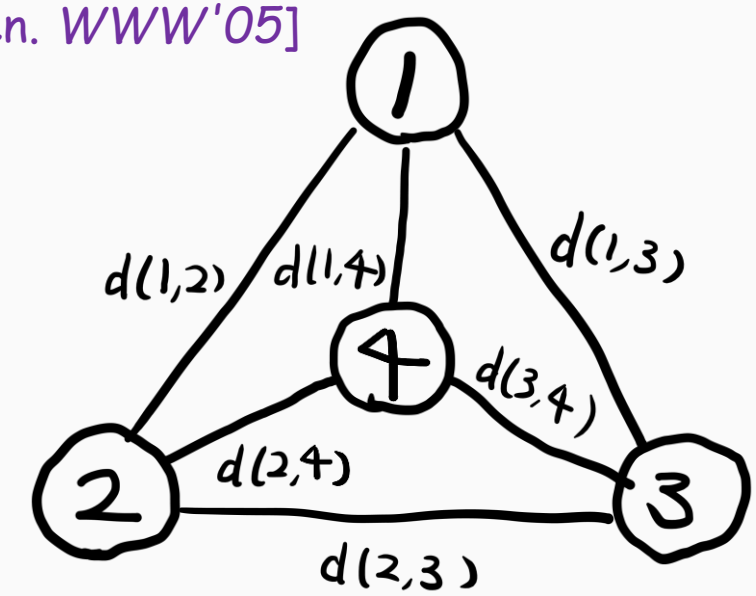
⚠ KEY: Define & optimize *appropriate* diversity objectives

Our target: Distance-based diversity

- Given distance d(i,j) between all item pairs i & j
- **Intra-list distance** (ILD) … Most popular distance-based objective [Smyth-McClave. *ICCBR'01*] [Ziegler-McNee-Konstan-Lausen. *WWW'05*]

  defined as average pairwise dist.

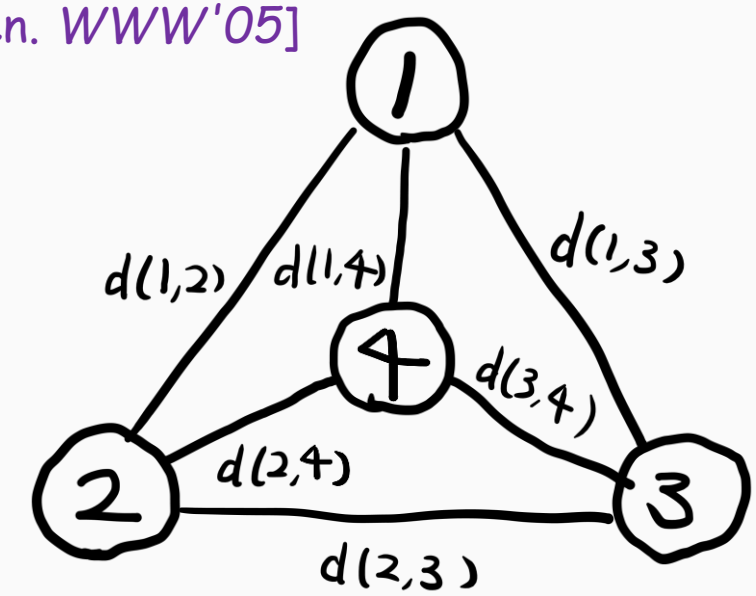
$$\mathrm{ILD}(S) \triangleq \frac{1}{\binom{|S|}{2}} \sum_{i \neq j \in S} d(i,j)$$



Our target: Distance-based diversity

- Given distance d(i,j) between all item pairs i & j
- **Intra-list distance** (ILD) ... Most popular distance-based objective
  [Smyth-McClave. *ICCBR'01*] [Ziegler-McNee-Konstan-Lausen. *WWW'05*]
  defined as average pairwise dist.

$$\mathrm{ILD}(S) \triangleq \frac{1}{\binom{|S|}{2}} \sum_{i \neq j \in S} d(i,j)$$



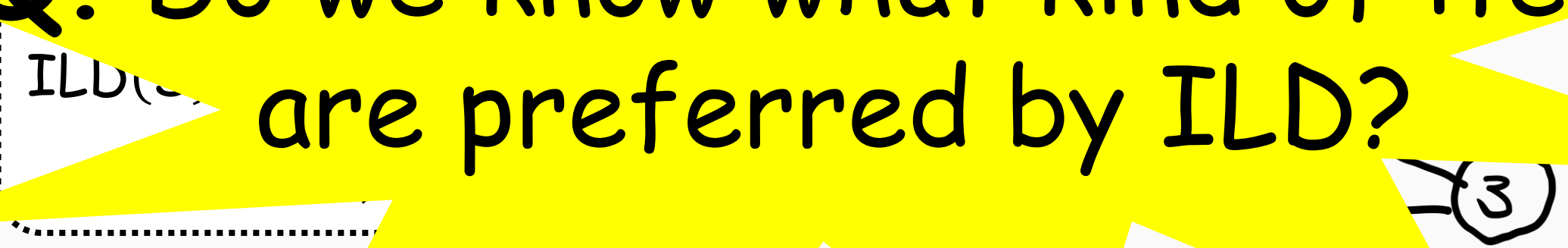
- 😋 Can use any distance metric depending on applications
- 😋 "Intuitive": Integrate pairwise distances
- 😋 Greedy heuristic works well [Birnbaum-Goldman. *Algorithmica'09*] [Ravi-Rosenkrantz-Tayi. *Oper. Res.'94*]

Our target: Distance-based diversity

- Given distance d(i,j) between all item pairs i & j
- **Int… list distance** (IL… Most po… stance-based objective
  [Smyth-M… …RR'01] [Zieg… …stan… …'05]

  defined as …

  ILD(… …



Q. Do we know what kind of items are preferred by ILD?

- 😋 Can use any …stance … …depen… …application
- 😋 "Intuitive": Integrate … …wise distan…s
- 😋 Greedy heuristic works …well

[Birnbaum-Goldman. *Algorithmica'09*]
[Ravi-Rosenkrantz-Tayi. *Oper. Res.'94*]

😜Quiz time!!!

1. Generate 1000 random points on ellipse↗
2. Select 128 points S by **maximizing ILD** (= average dist.)

   d(i,j) := Euclidean dist. btw. i & j

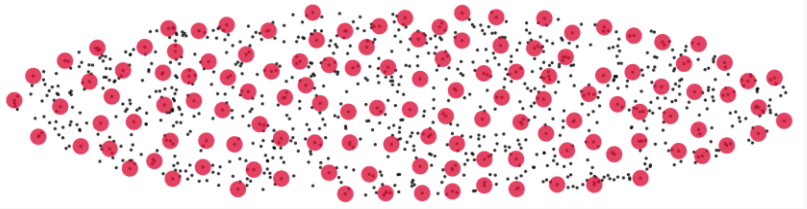
Introduction

😜Quiz time!!!

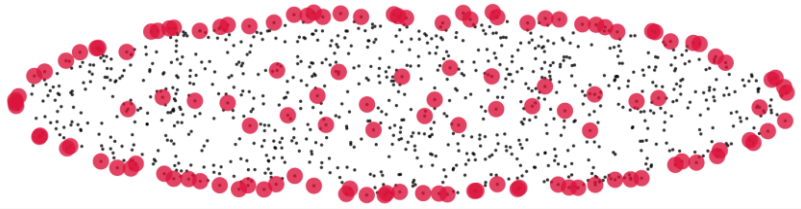
1. Generate 1000 random points on ellipse↩
2. Select 128 points S by **maximizing ILD** (= average dist.)

   d(i,j) := Euclidean dist. btw. i & j

Q. Which is S?



A



B

C

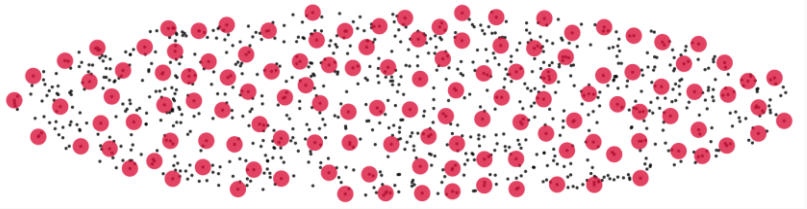
Introduction

😜Quiz time!!!

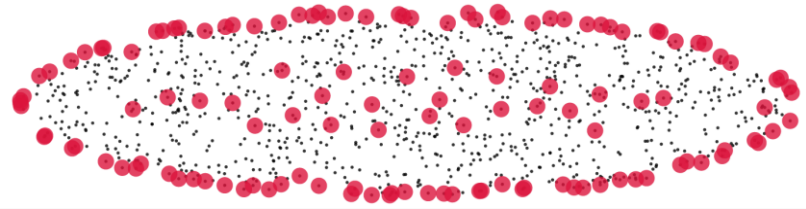
1. Generate 1000 random points on ellipse
2. Select 128 points S by **maximizing ILD** (= average dist.)

   d(i,j) := Euclidean dist. btw. i & j

Q. Which is S?



A



B

C

dispersion

= minimum dist.

[Erkut. *Eur. J. Oper. Res.*'90]
[Kuby. *Geographical Analysis*'87]

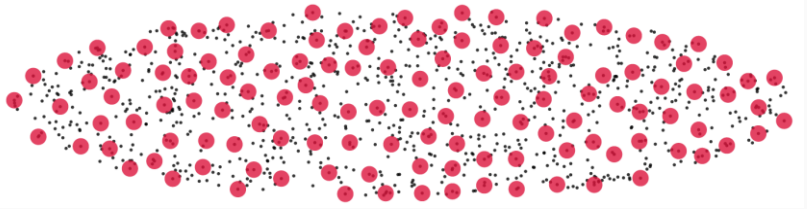
Introduction

Quiz time!!!

1. Generate 1000 random points on ellipse
2. Select 128 points S by **maximizing ILD** (= average dist.)

   d(i,j) := Euclidean dist. btw. i & j

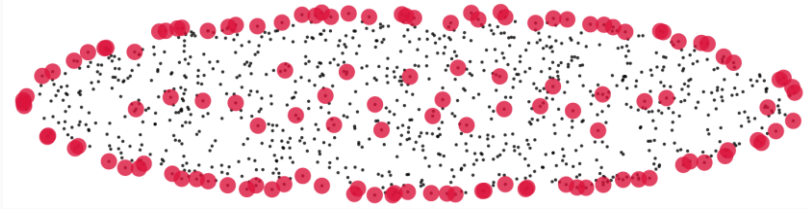
Q. Which is S?



A

dispersion

= minimum dist.



B

Gaussian ILD

= average Gaussian kernel

C

[Erkut. *Eur. J. Oper. Res.*'90]
[Kuby. *Geographical Analysis*'87]

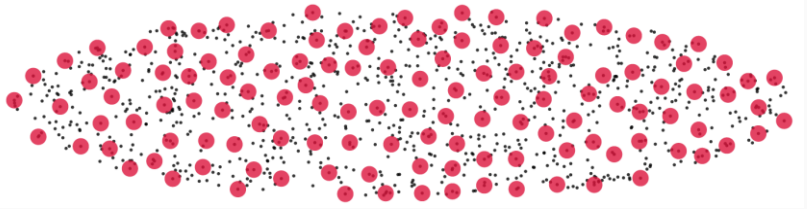
Introduction

Quiz time!!!

1. Generate 1000 random points on ellipse
2. Select 128 points S by **maximizing ILD** (= average dist.)

   d(i,j) := Euclidean dist. btw. i & j

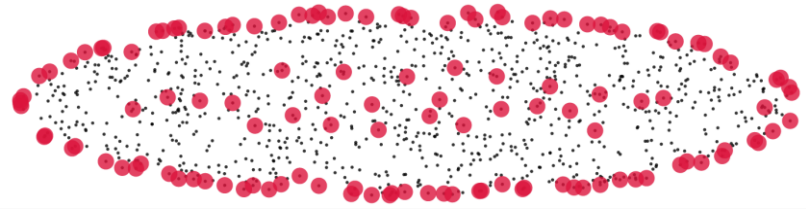
Q. Which is S?



A

dispersion

= minimum dist.



B

Gaussian ILD

= average Gaussian kernel

C

ILD

= average dist.

[Erkut. *Eur. J. Oper. Res.*'90]
[Kuby. *Geographical Analysis*'87]

Comparison analysis between ILD & dispersion

average dist. minimum dist.

In theory, we found...

- ILD may select **duplicate** items
- dispersion may **overlook distant** item pairs

Undesirable when recommending very few items.

$$\mathrm{ILD}(S) \triangleq \frac{1}{\binom{|S|}{2}} \sum_{i \neq j \in S} d(i,j)$$

$$\mathrm{disp}(S) \triangleq \min_{i \neq j \in S} d(i,j)$$

(OMITTED... see our paper)

- Experimentally verify the potential drawbacks
- Competitor interpolating btw. ILD and dispersion

Related work

Other diversity objectives

- Determinantal point processes [Borodin-Rains. *J. Stat. Phys.'05*] [Macchi. *Adv. Appl. Probab.'75*] [Kulesza-Taskar. *Found. Trends Mach. Learn.'12*]
  topical diversity
  [Agrawal-Gollapudi. *WSDM'09*] [Vargas-Baltrunas-Karatzoglou-Castells. *RecSys'14*]

Diversity enhancement algorithms

- Maximal marginal relevance (MMR) [Carbonell-Goldstein. *SIGIR'98*]
  local search [Yu-Lakshmanan-Amer-Yahia. *EDBT'09*]
  quadratic programming [Zhang-Hurley. *RecSys'08*]
  multi-objective optimization [Ribeiro-Lacerda-Veloso-Ziviani. *RecSys'12*]

  ⚠ Undesirable unless *appropriate* objective chosen

In information retrieval

- α-nDCG [Clarke-Kolla-Cormack-Vechtomova-Ashkan-Büttcher-MacKinnon. *SIGIR'08*]
  Intent-Aware [Agrawal-Gollapudi-Halverson-Ieong. *WSDM'09*]
  D# [Sakai-Song. *SIGIR'11*], αβ-nDCG [Parapar-Radlinski. *RecSys'21*]

  ⚠ They assume a distribution over the intent

Theoretical analysis

Our methodology

🎯 Quantify the **correlation** btw. two diversity objectives **f** & **g**

POLICY: Optimize w.r.t. **f**, evaluate w.r.t. **g**

1. Select S_f w.r.t **f** — **f**(S_f) maximized s.t. $|S_f|=k$
2. Evaluate S_f w.r.t. **g** — 🙄 Is **g**(S_f) also large?

If **g**$(S_f) > 0.999 \max_T$ **g**(T)

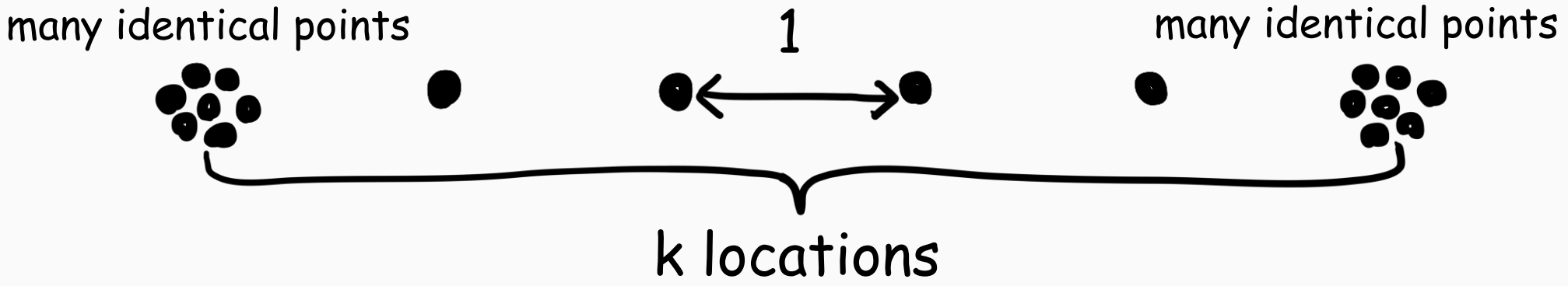
➜😋enhancing **f** also enhances **g**

If **g**$(S_f) < 0.001 \max_T$ **g**(T)

➜😖increasing **f** does NOT help improve **g**

Theoretical analysis

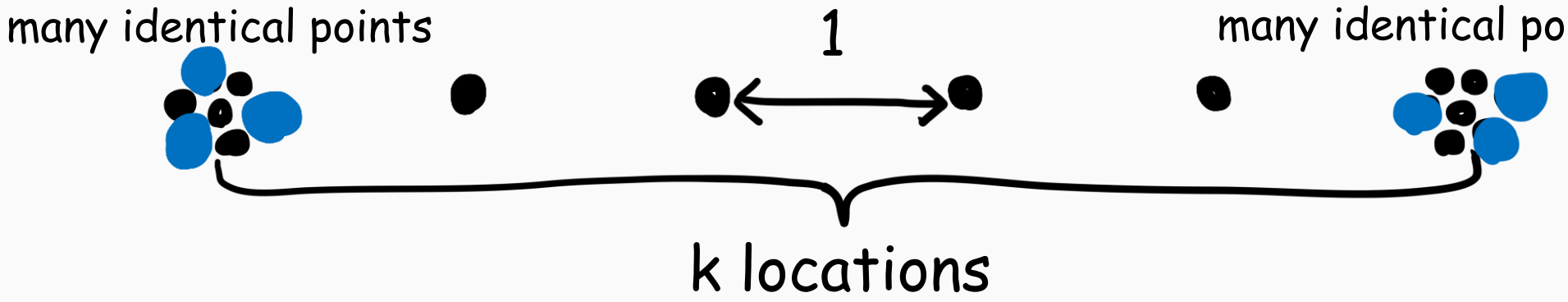
Does increasing **f**=ILD improve **g**=dispersion?



1. Select k items S_{ILD} w.r.t. ILD

Theoretical analysis

Does increasing f=ILD improve g=dispersion?

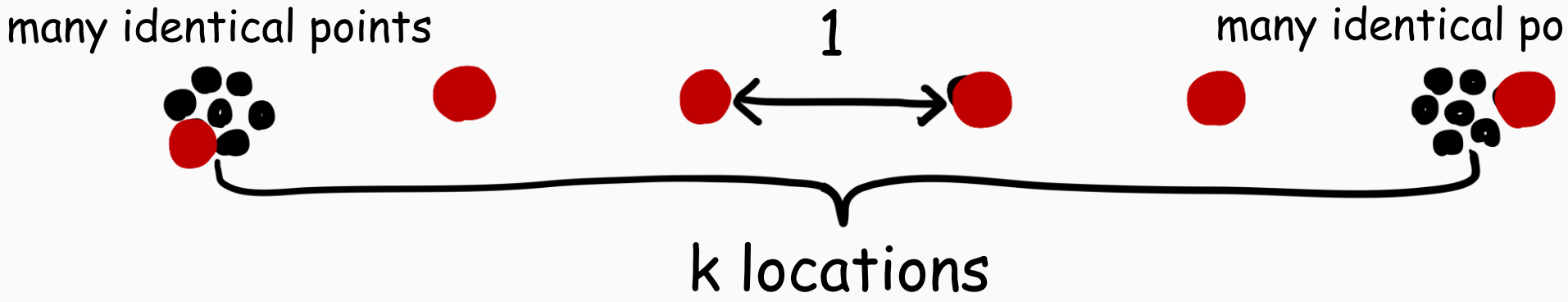


1. Select k items S_{ILD} w.r.t. ILD ➜ ILD prefers two ends (PROVABLY)
2. Evaluate S_{ILD} w.r.t. dispersion ➜ 😖 disp(S_{ILD}) = 0!

WHY? Select **duplicated items**

Theoretical analysis

Does increasing f=ILD improve g=dispersion?



1. Select k items S_{ILD} w.r.t. ILD ➜ ILD prefers two ends (PROVABLY)
2. Evaluate S_{ILD} w.r.t. dispersion ➜ 😖 $disp(S_{ILD}) = 0$!

WHY? Select **duplicated items**

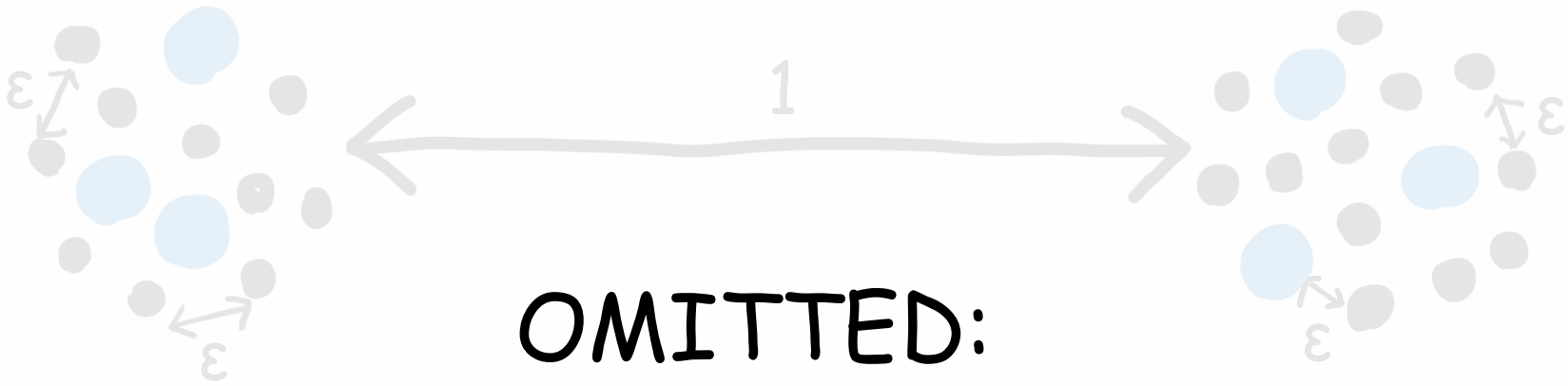
In fact... $\max_T disp(T) = 1$

CLAIM (Informal)

Even if S_{ILD} is optimal w.r.t. ILD, $disp(S_{ILD})$ can be 0

Theoretical analysis

Does enhancing **f**=disp enhances **g**=ILD?



OMITTED:

In a nutshell, we can do similar analysis

1. Select k items S_{disp} w.r.t. disp. ➔ disp. may be imbalanced
2. Evaluate S_{disp} w.r.t. ILD ➔ 😖 $ILD(S_{disp}) = \varepsilon \ll 1$

WHY? Can't **distinguish** small-ILD and large-ILD by disp.

In fact... $\max_T ILD(T) \approx 0.5$

CLAIM (Informal)

Even if S_{disp} is optimal w.r.t. dispersion,
$ILD(S_{disp})$ can be k times worse than $\max_T ILD(T)$

Conclusions: TAKEAWAY

Cons of ILD: May select (nearly) **duplicated** items

Cons of disp: May **overlook distant** item pairs

Pros of ILD: Select items in a **well-balanced** manner

Pros of disp: Selected items are **dispersed**