



Praktikum – Beating the Worst Case

Jean-Pierre von der Heydt und Marcus Wilhelm | 15.11.2023



Fragen zum Übungsblatt 0

Implementierung

- Wie seid ihr mit C++ klar gekommen?
- weitere Sprachen / Tools?
- habt ihr euer Auswertungs-setup automatisiert?

Aufgaben

- War klar, was die Aufgaben von euch verlangen?
- Wie schwer findet ihr die Aufgaben?

Auswertung:

Jonas, Philipp, Kilian: 400P
Sven, David: 400P (:
Victoria, Cedrico: 400P

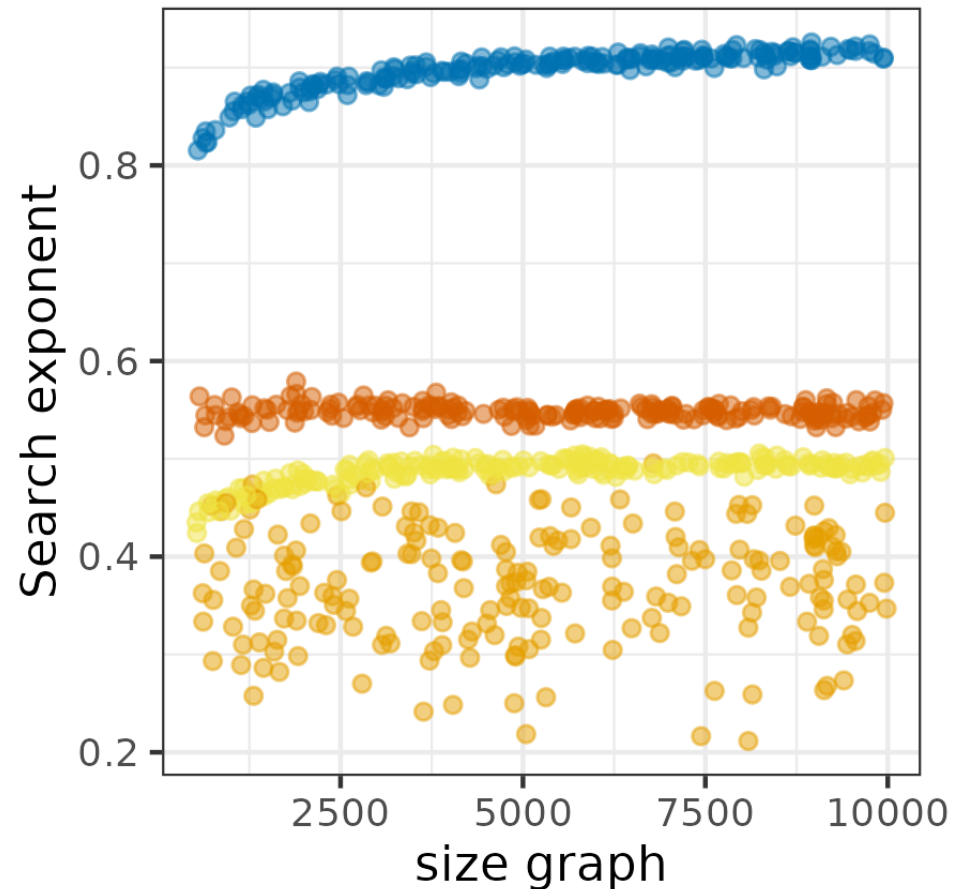
Ergebnisse

- Konntet ihr die Netzwerke deutlich unterscheiden?
- Sind euch Graphparameter eingefallen?
- Habt ihr eine Vermutung, wie die Graphen erzeugt wurden?



Lösung Übungsblatt 0

Bidirectional-BFS

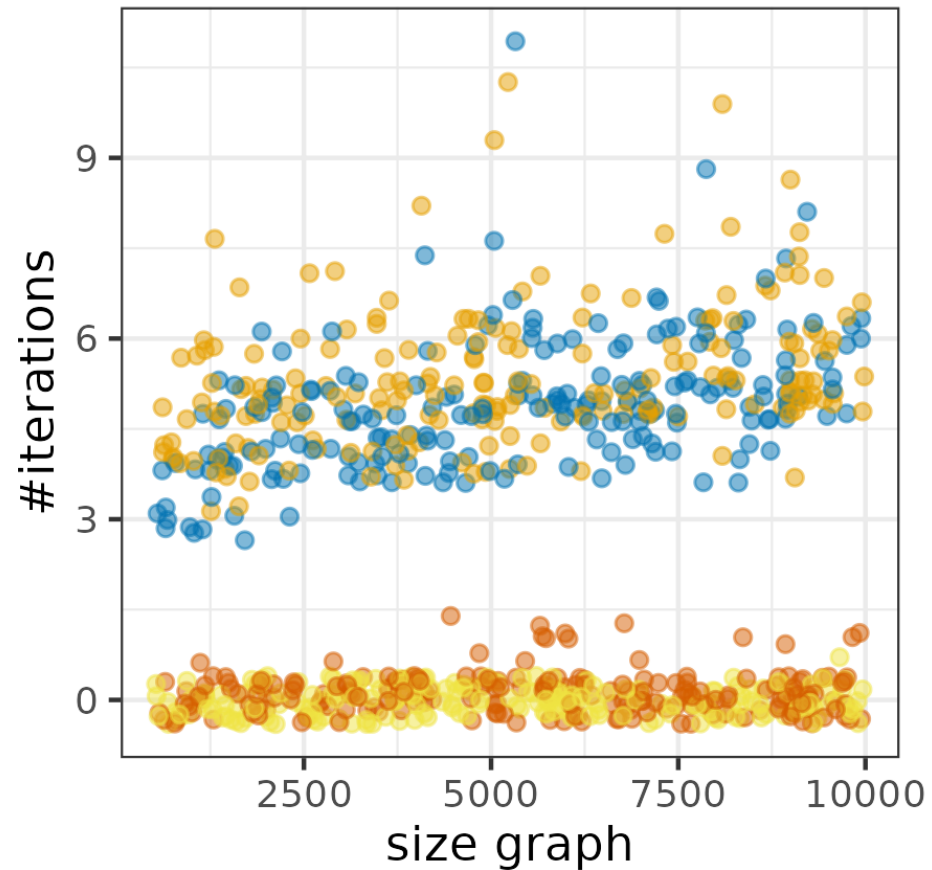


Für durchschnittliche
Suchraumgröße s ,
berechne x mit $s = m^x$
 $\Rightarrow x = \log_m(s)$



Lösung Übungsblatt 0

Vertex Cover #iterations



category

- 4
- 3
- 2
- 1

■ jeweils ~8min für das BFS/VC Experiment

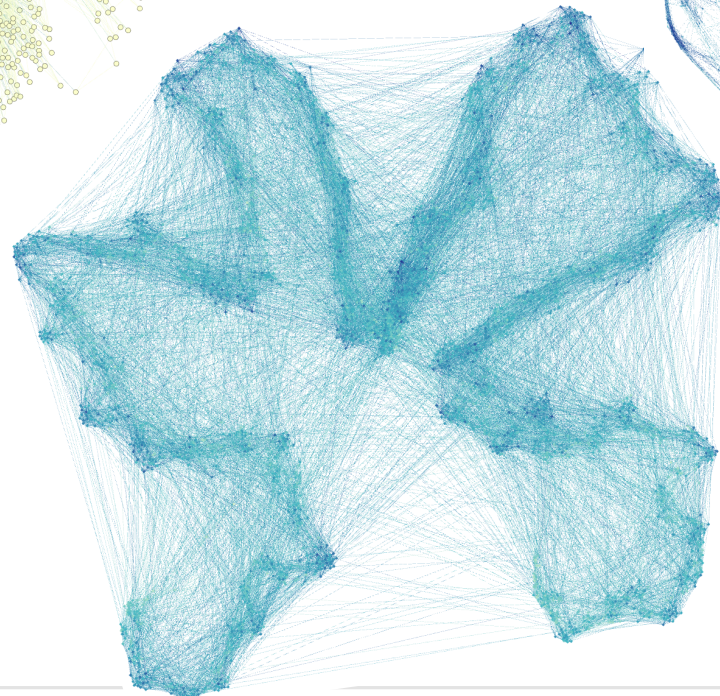
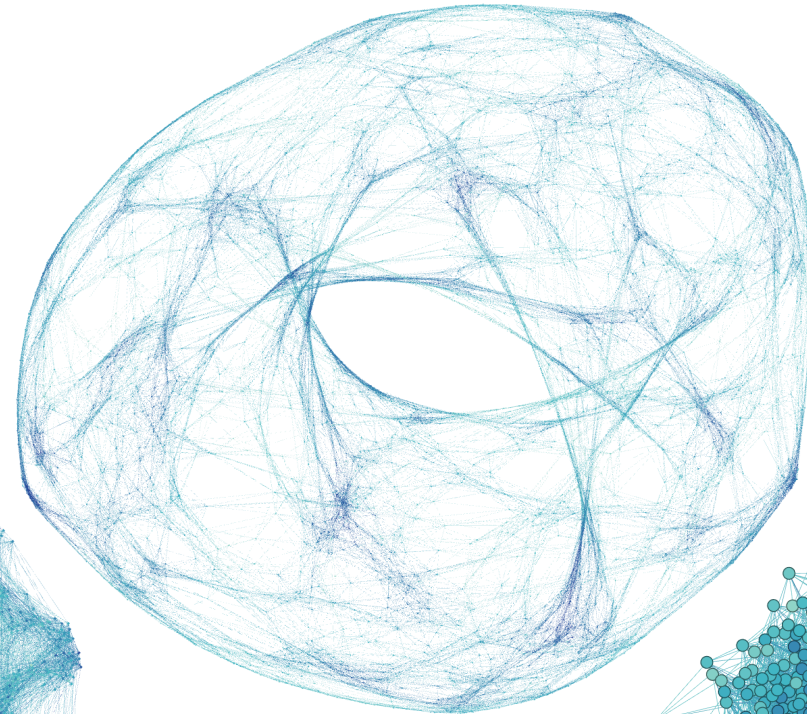
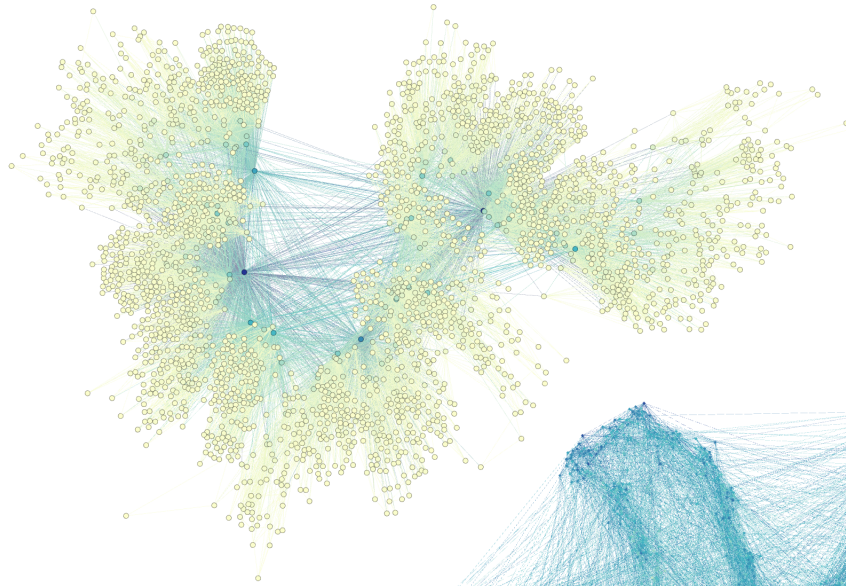


Wichtige Lösungsdetails

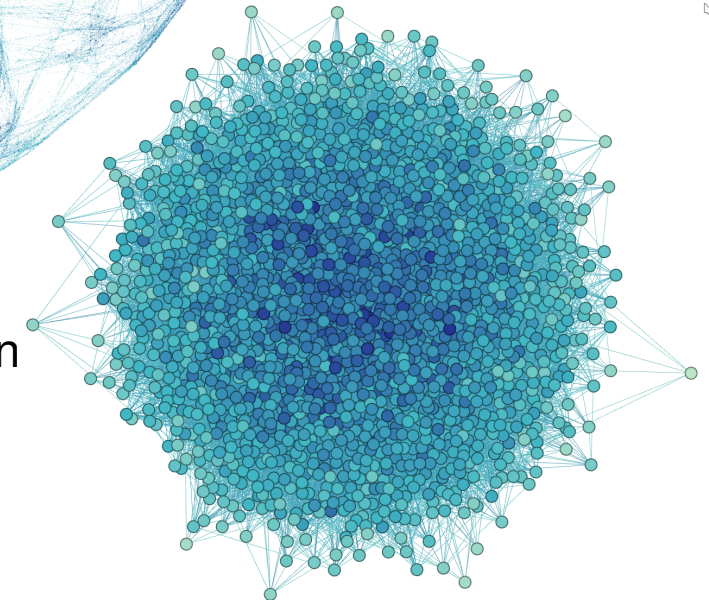
- Zeitmessungen können viel Rauschen erzeugen
 - Messung wiederholen und Mitteln
 - Suchraumgröße statt Zeit messen
- Messwerte Visualisieren
 - hilft Fehler oder Muster zu erkennen
- auf effiziente Implementierung achten
 - ~1 Mio. Operationen pro Sekunden
 - bei 10000 Knoten kein $\mathcal{O}(n^2)$
- hilfreich: automatisiertes Experiment Set-up



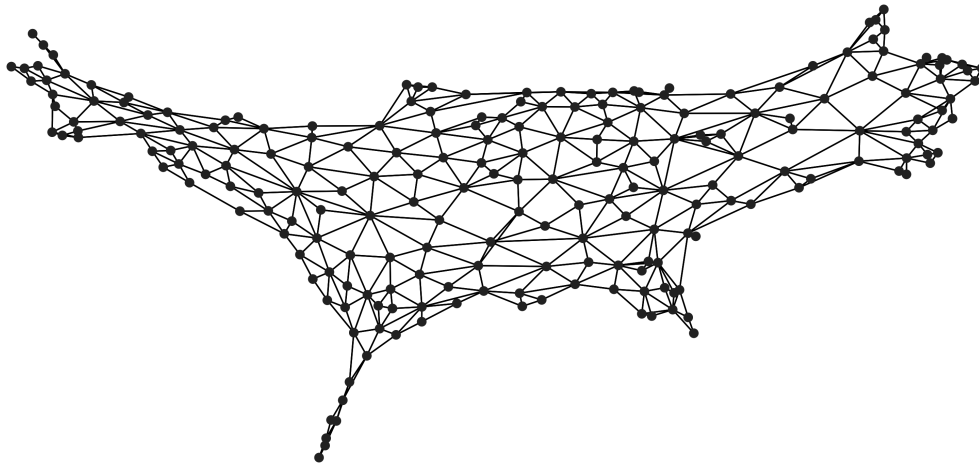
Netzwerke aus Datensatz



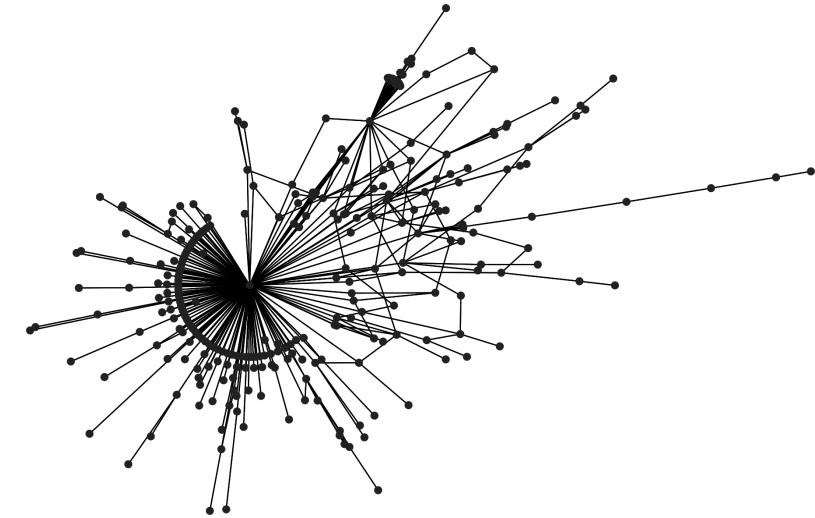
Motivation Blatt 1:
Unterschiede verstehen
& quantifizieren



Gradverteilung



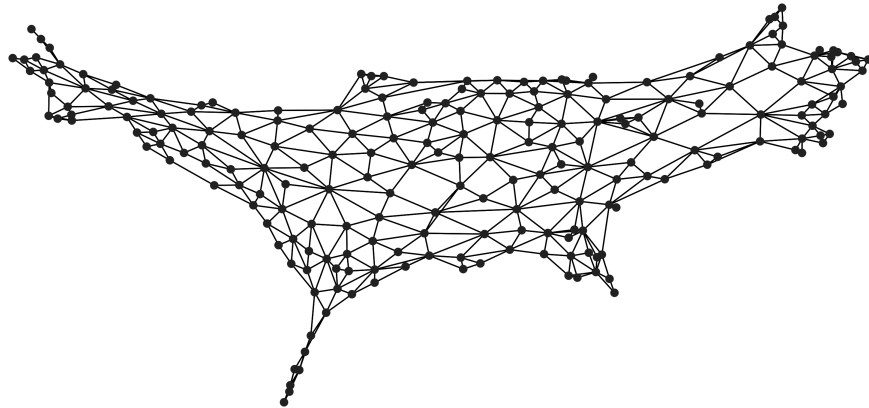
- Teil des US-Straßennetzes
- alle Knoten haben ähnlichen Grad



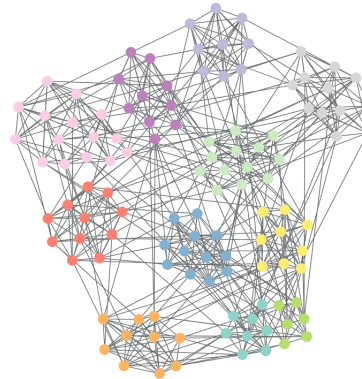
- Ausschnitt von Twitter Retweets
- zwei/drei extrem hochgradige Knoten

- Durchschnitts-/Maximalgrad nicht aussagekräftig genug
- konkrete Verteilung der Knoten ist relevant

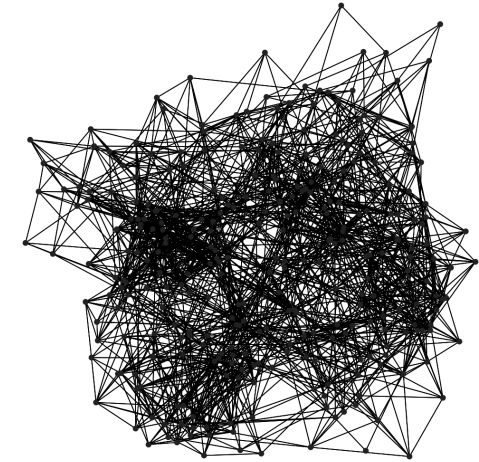
Lokalität



- Teil des US-Straßennetzes
- Knoten mit gemeinsamen Nachbarn sind oft selbst benachbart



- Spiele von College football teams
- Kanten entsprechen community structure



- Erdős–Rényi Graph
- Jede mögliche Kante existiert mit Wahrscheinlichkeit p (unabhängig)

- „Lokalität“, „Clustering“, „Geometrie“
- In vielen Echtwelt Netzwerken beobachtet (z.B. Straßen-, soziale Netzwerke)

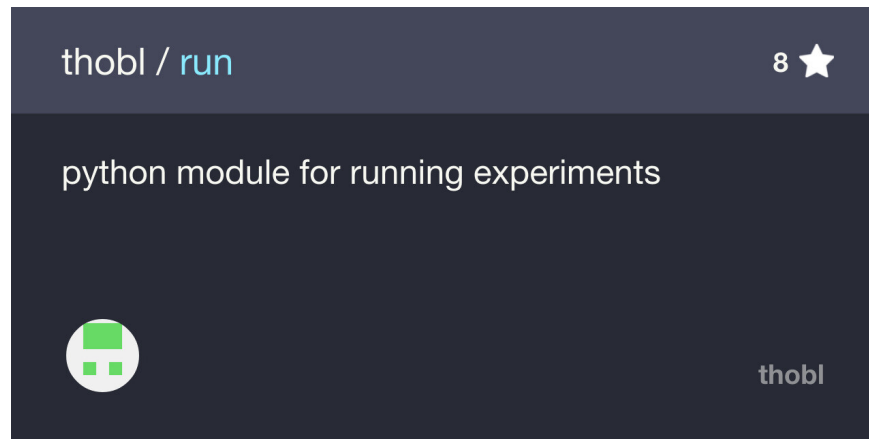
- keine „Lokalität“
- high *“temperature”*



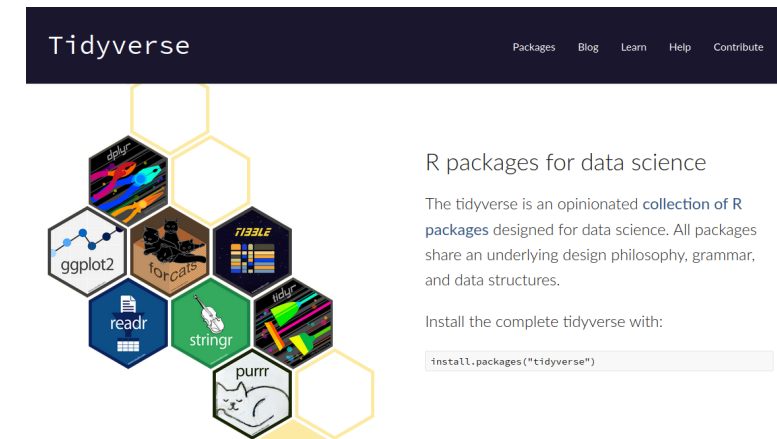
Workflow

Motivation: Erweiterbarkeit, Wiederholbarkeit

- Beispiel:
- neue Algorithmen
 - neue Eingaben
 - neue Messwerte



<https://github.com/thobl/run>



Plotting with R



Run

```
import run
from glob import glob
from os.path import glob

input_files = glob('input/*')

run.group("Stats")
run.add(
    "stats",
    "build/cli/bfs -s 100 --noheader [[input_files]]",
    {'input_files': input_files},
    stdout_file="output/bfs/stats.csv",
    header_command="echo 'stats' >>> stats.csv"
)

run.group("BiBFS")
run.add(
    "bibfs_cost",
    "build/cli/bfs -s 100 --noheader [[input_files]]",
    {'input_files': input_files},
    stdout_file=lambda args: f"output/bfs/{basename(args['input_files'])}.csv",
)

run.group("Postprocessing")
# add run to merge csvs

run.use_cores(4)
run.run()
```

```
marcus @ i11pcwilhelm in ~/work/teaching/praktikum_beating_the_worst_case_framework [18:41:44]
>>> python experiment.py BiBFS
Stats
├── stats
│   ├── bibfs_cost
│   └── merge_csvs
└── Postprocessing
    └── merge_csvs
        ├── stats
        └── bibfs_cost

running the experiments:
bibfs_cost: 5% | ██████████ | 34/630 [00:08<03:35, 2.77it/s]
```



Plotting with R

Was ist R?

- freie, open-source Programmiersprache
- primär für Statistik und Grafiken
- viele nützliche Pakete, z.B. ggplot2 und tidyr

ggplot2

- sehr umfangreiches und flexibles Paket für plotting
- basiert auf "Grammar of Graphics"
 - sehr strukturiert, hohe Abstraktion
 - komplexe Grafiken lassen sich leicht erstellen

Tidyverse, tidyr, dplyr

- Paket für Datentransformation
- strukturiert, elegant und einfach



Plotting with R – Example

```
# table bfs as before

bfs %>%
  mutate(
    speedup = cost_bi / cost_uni
  ) %>%
  ggplot(aes(x = cost_uni, y = cost_bi, color=speedup)) +
  geom_point() +
  xlab("Uni-directional cost") +
  ylab("Bi-directional cost")
```

Nützliche Ressourcen:

- R for Data Science, Chapter 1: Data visualization
- [Tutorial Website](#)
- Beispielcode anpassen
- Editor: RStudio, emacs
- [ggplot2 Cheat Sheet](#)

