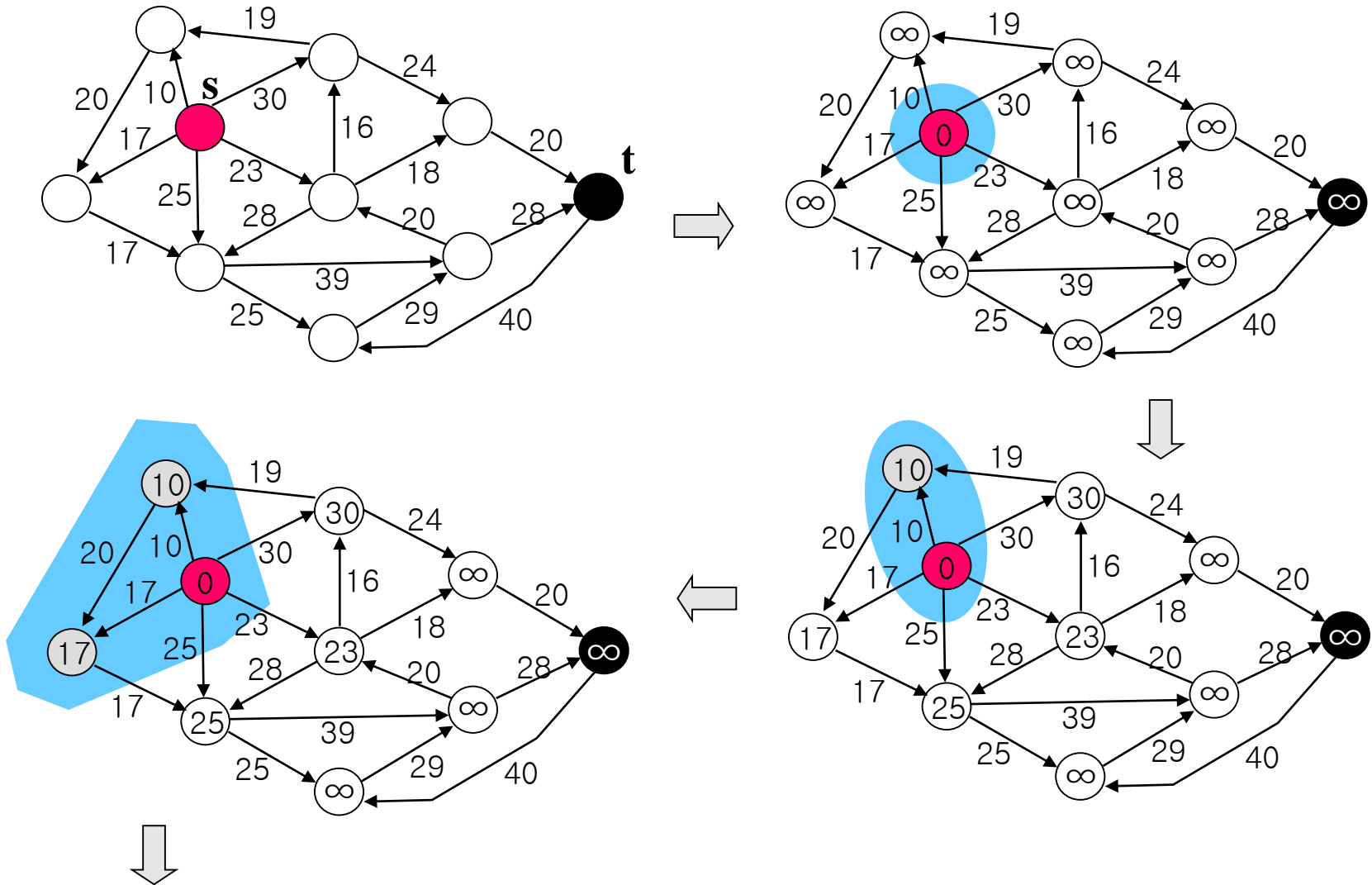


Homework #1

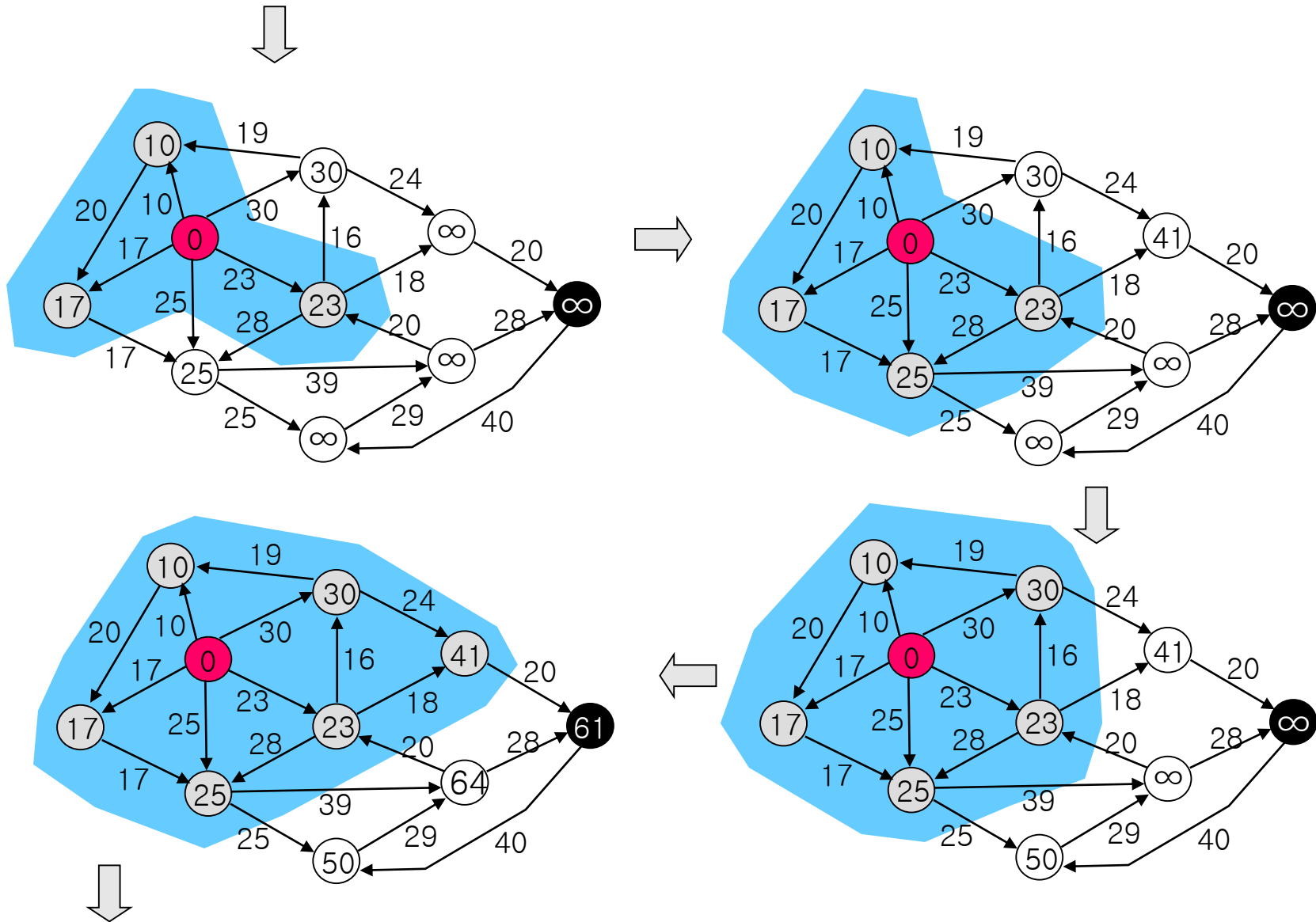
3. Dijkstra algorithm 의 비효율성

- Remind : Dijkstra algorithm
 - 시작점은 하나
 - 시작점으로부터 다른 모든 vertex에 이르는 최단 경로를 구한다 (목적점이 하나가 아니다)
 - 현실에서 사용하는 최단 경로는 목적지가 하나이다
 - 목적지 이외의 vertex를 계산하는 것은 비효율적
 - A* 알고리즘으로 해결 가능!

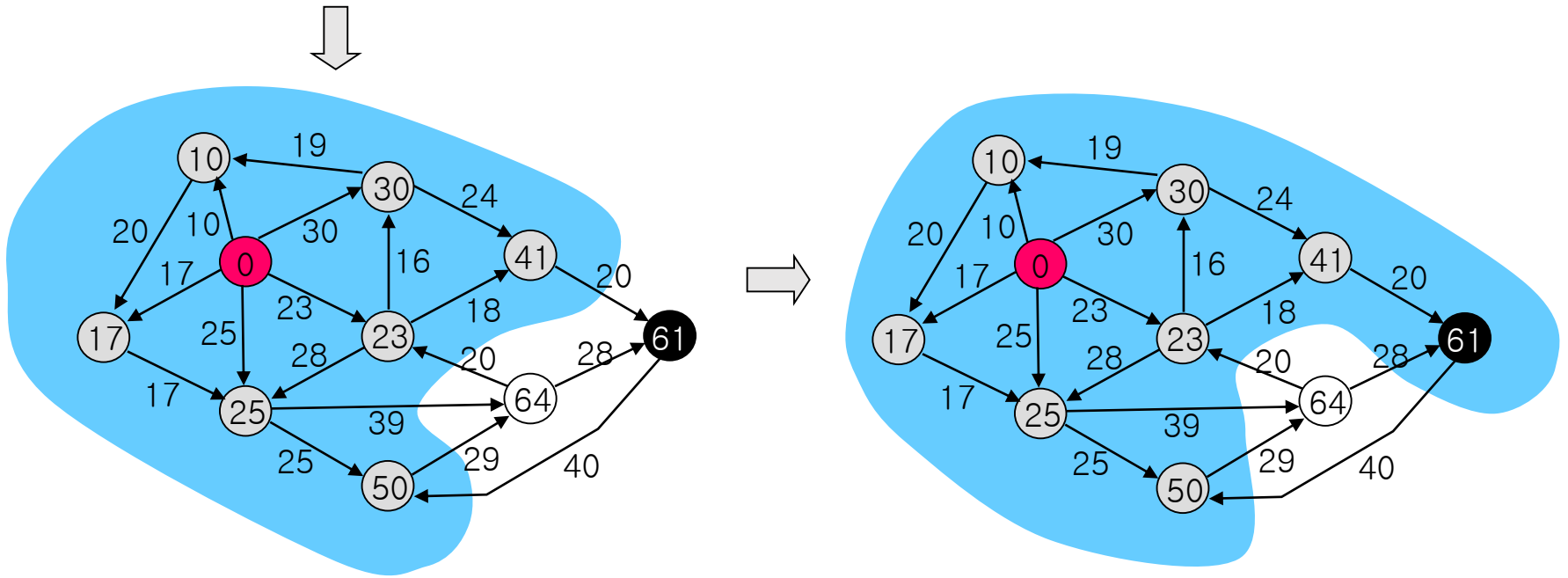
3. Dijkstra Algorithm의 작동 예



3.



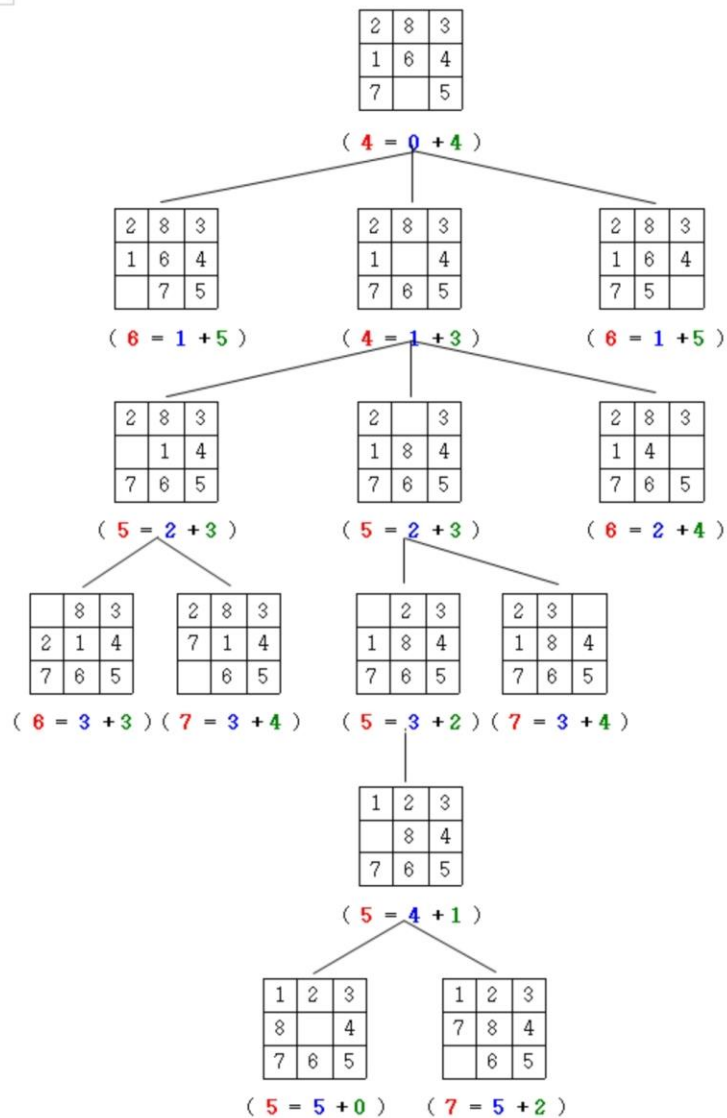
3.



3. A* 알고리즘

- Best-First-Search
 - 각 정점이 매력함수값 $g(x)$ 를 갖고 있다
 - 방문하지 않은 정점들 중 $g(x)$ 값이 가장 매력적인 것부터 방문한다
- A* algorithm은 best-first search에 목적점에 이르는 잔여추정거리를 고려하는 알고리즘이다
 - Vertex x 로부터 목적점에 이르는 잔여거리의 추정치 $h(x)$ 는 실제치보다 크면 안된다

3. A*-알고리즘의 대표적인 예: 8-Puzzle



1	2	3
8		4
7	6	5

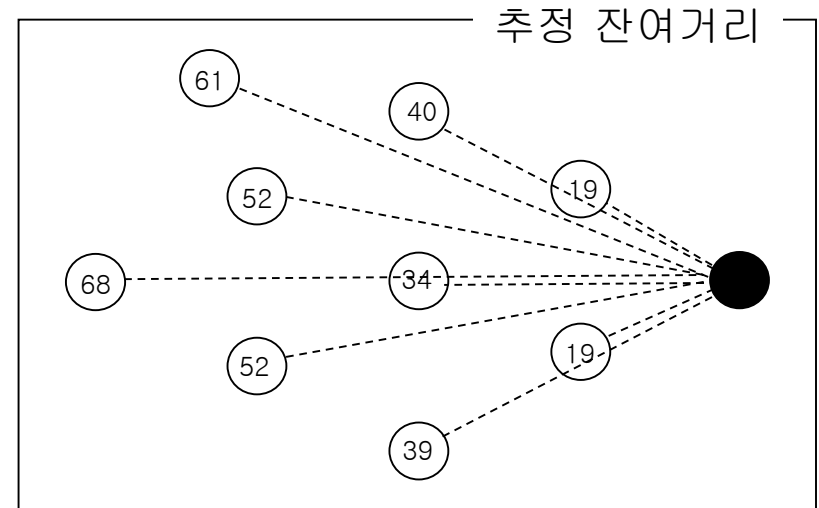
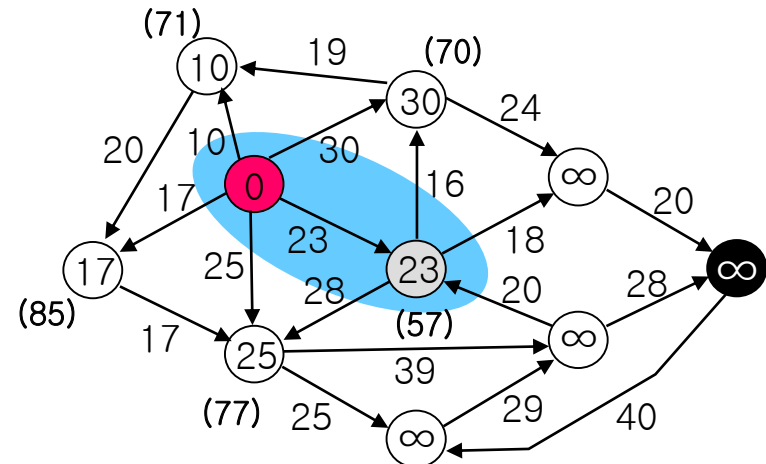
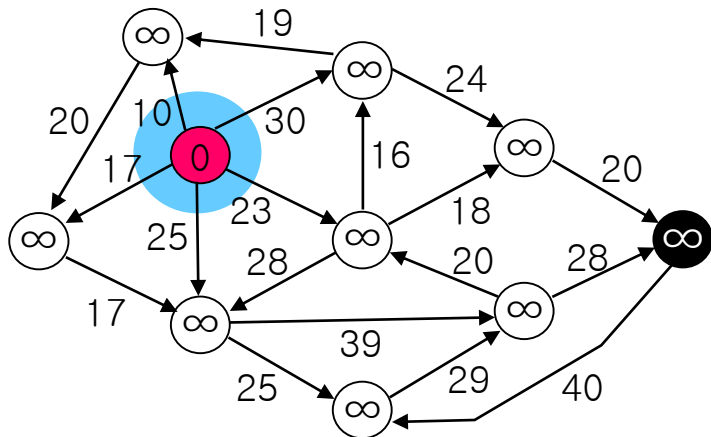
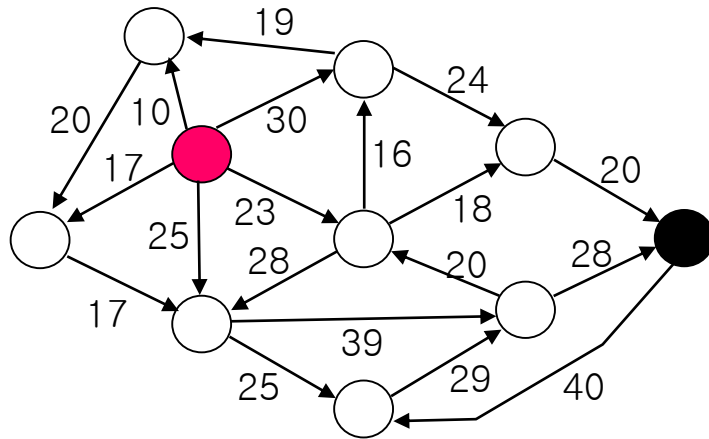
목표노드

$f(n) = g(n) + h(n)$

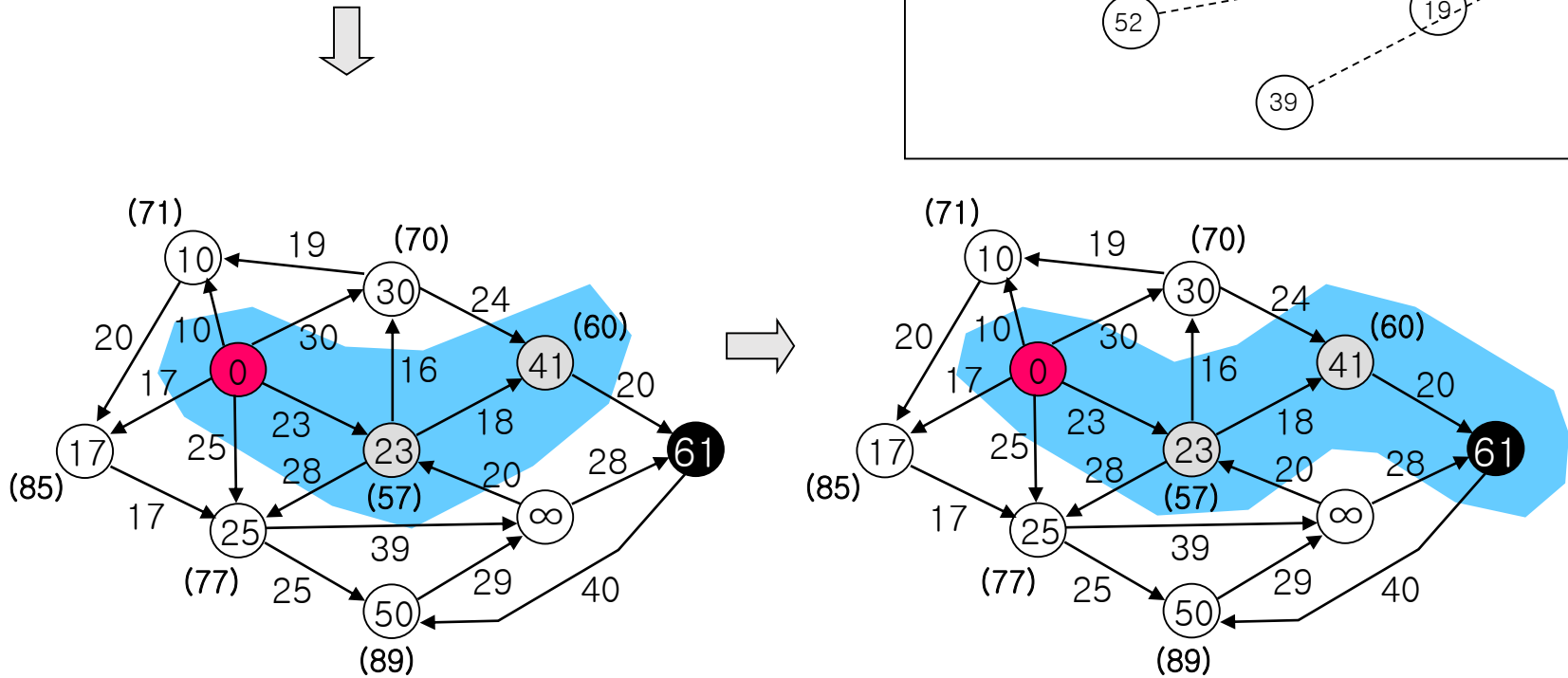
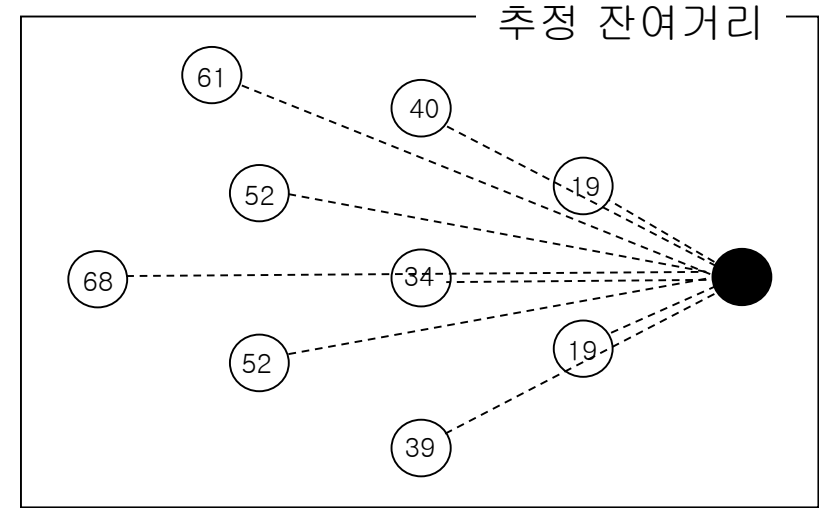
$g(n)$: 현재까지의 값, 즉 지금까지 움직인 횟수

$h(n)$: 앞으로 예상되는 값, 위에서는 제자리에 있지 않은 퍼즐의 수

3. A* Algorithm for Shortest Path



3.



✓ 추정잔여거리를 사용함으로써 탐색의 단계가 현저히 줄었다

과제

- Pyspark 를 이용하여 A* 알고리즘을 사용하는 Dijkstra 알고리즘을 구현
 - 구현 데이터로는 <http://www.dis.uniroma1.it/~challenge9>

- Goals
- Papers
- Competition
- Photos
- Organizers
- File formats
- Download**
- Workshop
- Mailing list
- Links

Download

- Challenge benchmarks
- Contributions by Challenge participants

Unless otherwise stated, the files available at this page contain data/software in the public domain and can be freely downloaded.

Challenge benchmarks


We have prepared a suite of benchmarks for the Challenge that includes synthetic input generators, real-world inputs, shortest path solvers, scripts for generating benchmark performance reports, and detailed documentation. The platform includes selected contributions by Challenge participants. For feedback, bug reports or comments please send mail to <goldberg at microsoft dot com> or <demetres at dis dot uniroma1 dot it>

- Download the Challenge 9 benchmarks - vs. 1.1 [ch9-1.1.tar.gz, 372 KB]
- View README file


The following table lists the 12 USA road networks that are part of the challenge core instances. Each graph comes in two versions: **physical distance** and **transit time** arc lengths. The node coordinates file is the same. For space reasons, this collection is not included in the experimental package, but it can be downloaded by the installer script.

Known issues: the data has numerous errors, in particular gaps in major highways and bridges. This may result in routes that are very different from real-life ones. One should take this into consideration when experimenting with the data.

Name	Description	# nodes	# arcs	Longitude	Latitude	Distance graph	Travel time graph	Coordinates
USA	Full USA	23,947,347	58,333,344	-	-	gr.gz file, 335 MB	gr.gz file, 342 MB	co.gz file, 218 MB
CTR	Central USA	14,081,816	34,292,496	[25.0; 50.0]	[79.0; 100.0]	gr.gz file, 195 MB	gr.gz file, 198 MB	co.gz file, 139 MB
W	Western USA	6,262,104	15,248,146	[27.0; 50.0]	[100.0; 130.0]	gr.gz file, 86 MB	gr.gz file, 88 MB	co.gz file, 57 MB
E	Eastern USA	3,598,623	8,778,114	[24.0; 50.0]	[-infy; 79.0]	gr.gz file, 49 MB	gr.gz file, 50 MB	co.gz file, 32 MB
LKS	Great Lakes	2,758,119	6,885,658	[41.0; 50.0]	[74.0; 93.0]	gr.gz file, 38 MB	gr.gz file, 39 MB	co.gz file, 24 MB
CAL	California and Nevada	1,890,815	4,657,742	[32.5; 42.0]	[114.0; 125.0]	gr.gz file, 26 MB	gr.gz file, 26 MB	co.gz file, 16 MB
NE	Northeast USA	1,524,453	3,897,636	[39.5; 43.0]	[-infy; 76.0]	gr.gz file, 21 MB	gr.gz file, 21 MB	co.gz file, 13 MB
NW	Northwest USA	1,207,945	2,840,208	[42.0; 50.0]	[116.0; 126.0]	gr.gz file, 15 MB	gr.gz file, 16 MB	co.gz file, 11 MB
FLA	Florida	1,070,376	2,712,798	[24.0; 31.0]	[79; 87.5]	gr.gz file, 14 MB	gr.gz file, 14 MB	co.gz file, 8.6 MB
COL	Colorado	435,666	1,057,066	[37.0; 41.0]	[102.0; 109.0]	gr.gz file, 5.5 MB	gr.gz file, 5.6 MB	co.gz file, 3.8 MB
BAY	San Francisco Bay Area	321,270	800,172	[37.0; 39.0]	[121; 123]	gr.gz file, 3.9 MB	gr.gz file, 4.0 MB	co.gz file, 2.5 MB
NY	New York City	264,346	733,846	[40.3; 41.3]	[73.5; 74.5]	gr.gz file, 3.5 MB	gr.gz file, 3.6 MB	co.gz file, 2.0 MB



9th DIMACS Implementation Challenge - Shortest Paths



데이터

- 도시 데이터 : 두 노드 사이의 양방향 경로로 되어있음. (최단 경로 구하는 데 문제없음)
- 도시의 위치 데이터 : 위도, 경도 표시 (추정 잔여거리로 활용 가능)

New York City, NY, USA

Latitude and longitude coordinates are: 40.730610, -73.935242.

USA-road-d.NY.gr - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

c 9th DIMACS Implementation Challenge: Shortest Paths

c <http://www.dis.uniroma1.it/~challenge9>

c TIGER/Line graph USA-road-d.NY

c

p sp 264346 733846

c graph contains 264346 nodes and 733846 arcs

c

a 1 2 803

a 2 1 803

a 3 4 158

a 4 3 158

a 5 6 774

a 6 5 774

a 7 8 1531

a 8 7 1531

a 9 10 1673

a 10 9 1673

a 9 11 1400

a 11 8 1400

USA-road-d.NY.co - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

c 9th DIMACS Implementation Challenge: Shortest Paths

c <http://www.dis.uniroma1.it/~challenge9>

c TIGER/Line nodes coords for graph USA-road-d.NY

c

p aux sp co 264346

c graph contains 264346 nodes

c

v 1 -73530767 41085396

v 2 -73530538 41086098

v 3 -73519366 41048796

v 4 -73519377 41048654

v 5 -73524567 41093796

v 6 -73525490 41093834

v 7 -73531927 41110484

v 8 -73530106 41110611

v 9 -73529341 41125895

v 10 -73529746 41127369

v 11 -73530583 41125051

v 12 -73529762 41085358

제출 안내

- 기한
 - 11/14일, 밤 12:00 까지 제출
 - 배점은 20점
 - Colab 파일로 KLAS 제출 (학번.ipynb)
 - Spark의 graphframe를 사용해야 함

Q & A

Thank you