

# Investigation of Causal Effects

Daniel Lindner, Simon Witzke

6/30/2020

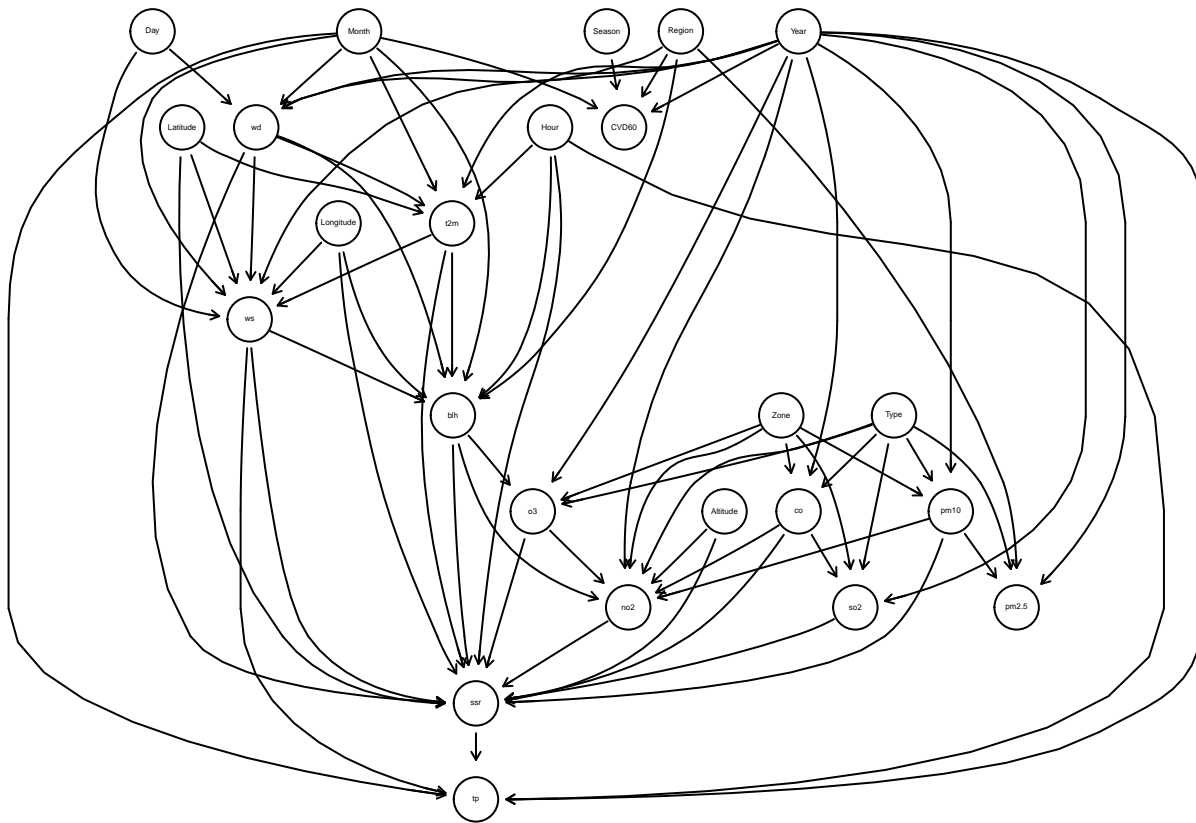
## Translation of learned bayesian network to igraph

```
sample.size <- 5e+06
sample <- generate.sample(sample.size)

clustered.data <- sample %>% base::sapply(function(x) if (!is.factor(x)) {
  factor(Ckmedian.id.dp(x, k = c(3, 5))$cluster)
} else x) %>% base::as.data.frame()

# TODO replace bn with our learned and saved model
adj.matrix.from.bngraph <- amat(bn)
igraph.from.bnmatrix <- graph_from_adjacency_matrix(adj.matrix.from.bngraph)
# TODO mark undirected edges as such
```

## Plot network



## Find causal effect

```
effect.igraph.from.bnmatrix <- causal.effect(y = "so2", x = "co",
  z = NULL, G = igraph.from.bnmatrix, expr = T)
```

$$\sum_{Zone, Type, Year} P(so2|Zone, Type, Year, co)P(Year)P(Type)P(Zone)$$

## Estimate causal effect strength

```
causal.effect.probabilites.so2.co <- function(data) {

  data %>% select(so2, co, Type, Zone, DateTime) %>% add_count(Zone) %>%
    mutate(prob_zone = n/nrow(data)) %>% select(-n) %>% add_count(Type) %>%
    mutate(prob_type = n/nrow(data)) %>% select(-n) %>% add_count(DateTime) %>%
    mutate(prob_datetime = n/nrow(data)) %>% select(-n) %>%
    group_by(Zone, Type, DateTime, co) %>% add_count(so2) %>%
    mutate(so2_cond_prob = n/nrow(data)) %>% select(-n) %>%
    mutate(total_prob = prob_zone * prob_type * prob_datetime *
      so2_cond_prob) %>% group_by(Zone, Type, DateTime, co,
      so2) %>% filter(n() == 1) %>% group_by(so2, co) %>% summarise(`P(so2|Type, Zone, DateTime, co)P
      # arrange(-'P(so2|Type, Zone, DateTime,
      # co)P(Type)P(Zone)P(DateTime)' )` %>%
    inner_join(data %>% group_by(co) %>% count(co) %>% mutate(`P(co)` = n/nrow(data)) %>%
      select(-n))
```

```

}

causal.effect.size.so2.co <- function(data) {
  sum(causal.effect.probabilitates.so2.co(data) %>% base::apply(MARGIN = 1,
    function(x) as.numeric(x[3]) * as.numeric(x[4]) * log(as.numeric(x[3])/sum(causal.effect.probab
      base::apply(MARGIN = 1, function(x) as.numeric(x[3]) *
        as.numeric(x[4])))))
}

xtable(causal.effect.probabilitates.so2.co(clustered.data))

```

% latex table generated in R 3.6.2 by xtable 1.8-4 package % Tue Jun 30 16:12:54 2020

|   | so2 | co | P(so2 Type, Zone, DateTime, co) | P(Type) | P(Zone) | P(DateTime) | P(co) |
|---|-----|----|---------------------------------|---------|---------|-------------|-------|
| 1 | 2   | 1  |                                 |         |         | 0.00        | 0.08  |
| 2 | 2   | 2  |                                 |         |         | 0.00        | 0.10  |
| 3 | 2   | 3  |                                 |         |         | 0.00        | 0.11  |
| 4 | 2   | 4  |                                 |         |         | 0.00        | 0.67  |
| 5 | 2   | 5  |                                 |         |         | 0.00        | 0.04  |
| 6 | 3   | 1  |                                 |         |         | 0.00        | 0.08  |

```
causal.effect.size.so2.co(clustered.data)
```

```
[1] 5.369082e-11
```